THE SEARCH FOR ARCHAEAL VIRUSES IN HIGH TEMPERATURE ACIDIC ENVIRONMENTS AND CHARACTERIZATION OF *SULFOLOBUS TURRETED ICOSAHEDRAL VIRUS* (STIV)

by

George Ernest Rice III

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

August 2006

APPROVAL

of a dissertation submitted by

George Ernest Rice III

This dissertation has been read by each member of the dissertation Committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Mark Young

Approved for the Department of Microbiology

Dr. Tim Ford

Approved for the College of Division of Graduate Education

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microfilm along with the non-exclusive right to reproduce and distribute my abstract in any form in whole or in part."

George Ernest Rice III

23 August, 2006

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF TABLES

Table Page

LIST OF FIGURES

Figures Page

LIST OF FIGURES – CONTINUED

Figure Page

ABSTRACT

Viruses of extreme thermophiles are of great interest because they can serve as model systems for understanding biochemical molecular nuances required for life at high temperatures. This two part body of work first reports the discovery and isolation of viruses and virus-like particles from extreme thermal acidic environments (70-92°C, pH 1.0-4.5) in Yellowstone National Park (YNP), and secondly details the characterization of one of these viruses that possesses a capsid structural motif that is found in at least two other families of viruses inhabiting the other two domains of life (Bacteria and Eukarya). This is of particular interest because it is the first example of a predicted but previously undocumented structural relationship between any entities (living or non) that span all three domains of life. The implications of this reported connection lend credence to the theory that there was a common viral ancestor, or ancestors, that predate the division of life into the three separate currently accepted domains.

CHAPTER 1

THERMOACIDOPHILLIC ARCHAEAL VIRUSES: BACKGROUND AND SCOPE OF THIS RESEARCH

The objectives of this research are to carry out a survey of the thermophilic viruses present in the thermal features in Yellowstone National Park (YNP), and to perform a preliminary characterization of one of the viruses identified in that survey. Previous studies had shown that thermophilic viruses had been found in thermal features such as mud pots, fumaroles, geysers, and hot pools in other parts of the world including New Zealand, Iceland, Italy, and Japan (Arnold et al. 1999b; Zillig et al. 1994; Zillig et al. 1996). However, no such survey had been carried out in YNP. The aforementioned studies employed a host-dependant approach (culturing of *Sulfolobus*) for the isolation and characterization of viruses from these environments.

This chapter introduces the concept of thermophilic prokaryotic organisms and discusses the delineation of the third domain of life, Archaea. Included are the morphological, metabolic, and replicative differences as well as similarities between Archaea and the other two domains of life (Bacteria and Eukarya). The chapter also covers bio-applications of thermophiles as well as an overview of the thermophilic Crenarchaeal viruses that have been characterized to date.

Major Questions and Hypothesis

Major questions this research effort proposes to address.

1. Can we detect and isolate viruses in and from the thermal features of Yellowstone National Park?
2. What is the best approach to detect and isolate thermal viruses in YNP if they are present?
3. If there are viruses present in this environment, are they similar in morphology and genome content to viruses found in other thermal areas around the world?
4. If we do find thermal viruses in YNP is it possible to characterize a previously uncharacterized morphotype, and if so what might it tell us about life in extreme environments and the relationships of these viruses to one another?

We hypothesize that there are viruses present in the thermal acidic environments of Yellowstone and that the isolation and characterization of these viruses will increase the understanding of thermal virology in general.

Thermophiles

Thermophiles are defined as the members of a group of organisms that routinely exist close to the upper temperature limit of life for that group, be it Archaea, Bacteria, or Eukarya. As outlined by Thomas Brock, this is an arbitrary range that varies with the group (Brock 1995). Early research on thermophiles arose from two distinct fields of

bacteriology. The first was the development of conventional bacteriological culture procedures designed during the study of bacteria that survive the elevated temperatures employed to preserve food in the canning industry. The second type of research on thermophiles arose from field studies of organisms that live in geothermal habitats. The early stages of this discipline were primarily focused on phototrophs (organisms that use light as their main source of energy), which attracted attention due to the visible colors of the organisms living in hot springs. This early work did not lead to the discovery of any organisms that are currently considered thermophiles due to the relatively low temperature maximums under study (65°C for phototrophs and ~50°C for the bacteria which were of concern in the food industry).

Early biologists, previous to the prototroph and food industry work, reported observations of organisms living in geothermal waters up to 80°C. These studies go as far back as the Roman era with Pliny the Elder (Cohen 1995). The botanist Ferdinand Cohn (Kirk and Gruber 2005) and 19th century biochemist Hoppe-Seyler (Noyer-Weidner and Schaffner 1995). So, early on biologists were aware of high temperature organisms from observation of their existence in geothermal waters up to 80°C but their existence was forgotten due to the extensive work of food bacteriologists where the standard maximum culturing temperature became 55°C. The consequence of this was that by the 1950s the definition of a thermophile was as an organism that grew around a 55ºC temperature maximum. There were no studies of higher temperature thermophiles in their natural habitats.

Yellowstone National Park has the highest concentration of thermal features in the world, and work has been performed there observing the organisms that occupy these extreme habitat niches (Brock 1995). In 1888, the geologist W.H. Weed was the first to study the thermal biology of this area out of an interest in the role that organisms might be playing in silica and travertine depositions. He was the first to show that the color found in outflow channels was not always the result of the underlying mineralogy. In 1903 W.A. Setchell published a paper entitled, "The Upper Temperature Limit for Life," in which he made a distinction between phototrophic life, which has a temperature limit of about 75°C, and non-phototrophic bacteria, which he found living in abundance at 82-89°C.

Over the past 30 years the definition of what constitutes an extreme thermophile or hyperthermophile has changed. *Thermus aquaticus*, which lives optimally at 79°C, was characterized as an extreme thermophile by Brock in 1969 (Brock 1997). This temperature boundary has crept up over the years with continued research and although the precise definition of an extreme thermophile is still subject to debate, it is generally accepted that thermophiles are those organisms that live optimally above 60°C and extreme, or hyper, thermophiles are those that do the same above 80°C.

Domain Archaea

In the 1980s Carl Woese's interpretation of molecular data resultant from revolutionary techniques employed to phylogeneticly classify prokaryotic organisms based on their 16S ribosomal DNA sequences led him the concept that a portion of these organisms (especially those from thermal acidic environments) were different enough to

warrant a new name (Woese et al. 1990). These unique prokaryotes, which he called archaeabacteria, were actually different enough on a molecular level that they warranted the formation of a whole new domain of life. After some contentious debate in the scientific community, Archaea is now firmly established as a third domain of life (Achenbach-Richter et al. 1987).

Before 1980 only methanogenic bacteria were considered archaeabacteria, but in 1980, through the advent of molecular classification of 16S ribosomal RNA, it was established that *Thermolasma* and *Sulfolobus* also belonged to this new domain. Soon after, Wolfram Zillig, Karl Stetter and colleagues began extensive studies of thermophilic archaea and their extra-chromosomal elements. This work led to the development of culturing techniques for organisms that live above 100°C (Stetter 1995) facilitating the discovery of many extreme thermophiles. The following realization that archaea are well represented in the realm of extreme thermophiles has stimulated interest in the field. For instance, it has been reported that pelagic archaea made up more than 30% of the prokaryotic biomass in Antarctic costal surface waters (Wang et al. 2004) and more than 30 species of archaea have been identified by ribo-typing from a single thermal feature in YNP alone (Barns et al. 1994).

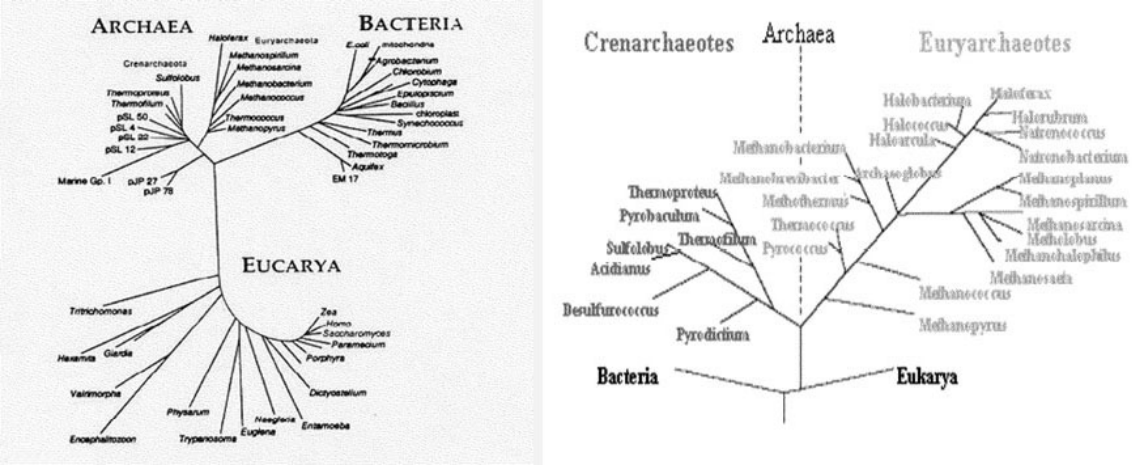

Figure 1.1 - Left; three domain tree of life constructed from 16s rDNA sequences (Pace, 1997), Right; the two kingdoms of Archaea–-Euryarchaeota and Crenarchaeota.

Archaea are well represented in the niches that are defined by extreme environments. These environments can include conditions such as high salinity (above 1M); high pH (above 9) or low pH (below 4); temperatures above 100°C; extreme pressure; or any combination of these. The Archaeal domain is divided into the Euryarchaeota and the Crenarchaeota kingdoms. Two other kingdoms have been proposed for addition to this domain. They are Korarchaeota, which branch near the root of the archaeal tree but have been extremely hard to culture, and Nanoarchaeota. Members of the proposed Nanoarchaeota are small symbiotic organisms (~100 nm), not much larger than a virus (one tenth the size of the associated cell) found only in conjunction with other archaea (Waters et al. 2003). The Euryarchaeota kingdom is comprised mostly of methanogenic and halophilic organisms that are distinct from the Chrenarchaeotes based on their 16s rDNA sequences (Garrity 2001). The bulk of Crenarchaeotes studied to date are from terrestrial and marine environments and are hyperthermophilic in nature (Barns et al. 1996). There are species of Chrenarchaeotes

that are found in non-thermal marine environments. They are planktonic in nature and occur in large numbers (DeLong 2001). The three orders of the Chrenarchaeal kingdom that have been studied the most to date are the *Thermoproteales* (anaerobic sulfur-reducers), *Sulfolobales* (aerobic sulfur-oxidizers), and the *Desulfurococcales* containing the highest temperature archaea to date from a deep marine thermal vent. Some characterized *Desulfurococcales* have optimum growth temperatures greater than 100°C

Although archaea are prokaryotic in nature and bear a morphological resemblance to bacteria, they are as phylogeneticly distinct from that domain as they are from eukaryotes. For instance it has recently become apparent that the DNA and RNA replication mechanisms of archaea have more in common with eukaryotic systems than those of bacteria (Bell and Dutta 2002). The relative simplistic nature of the archaeal genetic replication apparatus makes an attractive model system for study of the central mechanisms of eukaryotic DNA replication. Biochemical studies and genome sequencing have shown specifically that the archaeal basal transcription machinery is more homologous to the eukaryal RNA polymerase (RNAP) II core machinery than to that of bacteria (Soppa 1999). This is evidence that the RNAPII machinery evolved prior to the divergence of the archaeal and eukaryal domains but after those two domains had diverged from Bacteria (Bell et al. 1999).

To date three groups of transcription associated proteins have been identified in Archaea (Wang et al. 2004). One group is significantly homologous to eukaryotic transcription proteins; another with those found in Bacteria. A third group, showing similarity to those found both in Eukaryotes and Bacteria (Ouhammouch et al. 2003), has

shown that archaea have a eukaryotic-like apparatus for transcription that incorporates a bacterial-type regulator that is involved in the recruitment of the TATA-binding protein. Several homologs of bacterial transcriptional factors are involved, such as a metal-dependent repressor MDR1 (Bell and Dutta 2002); a leucine-responsive regulatory protein Lrp (Dahlke and Thomm 2002); a heat shock regulator Phr (Vierke et al. 2003); and the transcriptional regulator of the mal operons Trm (Lee et al. 2003). All of these transcriptional factors have been characterized in Archaea, so it appears that archaeal transcriptional apparatuses have maintained their own mechanistic patterns while showing similarities to those of eukaryotes and bacteria (Kyrpides and Ouzounis 1999).

Two distinct classes of translation initiation mechanisms have been identified in archaea—leadered and leaderless translation. Leadered translation, which is similar to that of bacteria, is associated with the expression of internal genes of archaeal operons (Tolstrup et al. 2000). It involves the binding of 16S rRNA to the Shine-Dalgarno sequence of the mRNA. Genes that are translated in this manner have a G-rich region in their 5' flanking region that is complimentary to the Shine-Dalgarno consensus sequence. Leaderless translation involves scanning along the transcript for the first start codon (Kozak 1999), and isolated genes as well as the leading genes in many archaeal operons are involved in this type of eukaryotic-like translation system (Tolstrup et al. 2000).

Sulfolobus and Archaeal Metabolism

Sulfolobus solfataricus is a thermoacidophilic crenarchaeon that thrives at between 70°C and 90°C and in a pH range of 2-4. It has been a model crenarchaeal organism since its isolation in the early 1980s. This is largely due to the ease with which

it can be collected in the field and cultured in the lab. The *S. solfataricus* genome was completed in 2001 (She et al. 2001) furthering its popularity as an interesting source for novel thermostable enzymes as well as other biochemical adaptations required for life at an elevated temperature and reduced pH. Currently, 1941, (53%) of its putative genes in The Institute for Genomic Research (TIGR) comprehensive microbial resources (CMR) have no known function. Of the original open reading frames (ORFs) identified in the *S. solfataricus* genome, 40% of the genes are archaeal in nature, 12% are bacterial, and 2.3% are shared specifically with eukaryotes (Peterson et al. 2001).

The availability of the genome sequences of *S. solfataricus, Sulfolobus tokodaii* (Kawarabayasi et al. 2001b), and *Sulfolobus acidocaldarius* (Chen et al. *unpublished*; http://dac.molbio.ku.dk) have allowed for the identification of genes encoding proteins that have been identified in metabolic proteomic studies on *Sulfolobus* species. With these results, a putative reconstruction of the central carbon metabolic pathways was preformed (Snijders et al. 2006).

In the metabolic pathways of all three domains of life polysaccharides are major sources of carbon supporting heterotrophic growth (Verhees et al. 2003). Utilization involves extra cellular and intracellular hydrolysis of these polysaccharides into hexoses and pentoses. These monosaccharides are then oxidized via a well-conserved set of central metabolic pathways. The Embden-Meyerhof (EM) pathway is the general route for glucose degradation in all domains of life. Some organisms degrade glucose via an alternate pathway called the Entner-Ddoudoroff (ED) pathway, and others are capable of bypasses in sugar degradation such as the oxidative pentose phosphate pathway.

Established biosynthetic pathways include gluconeogenesis and the non-oxidative pentose phosphate pathway (Verhees et al. 2003).

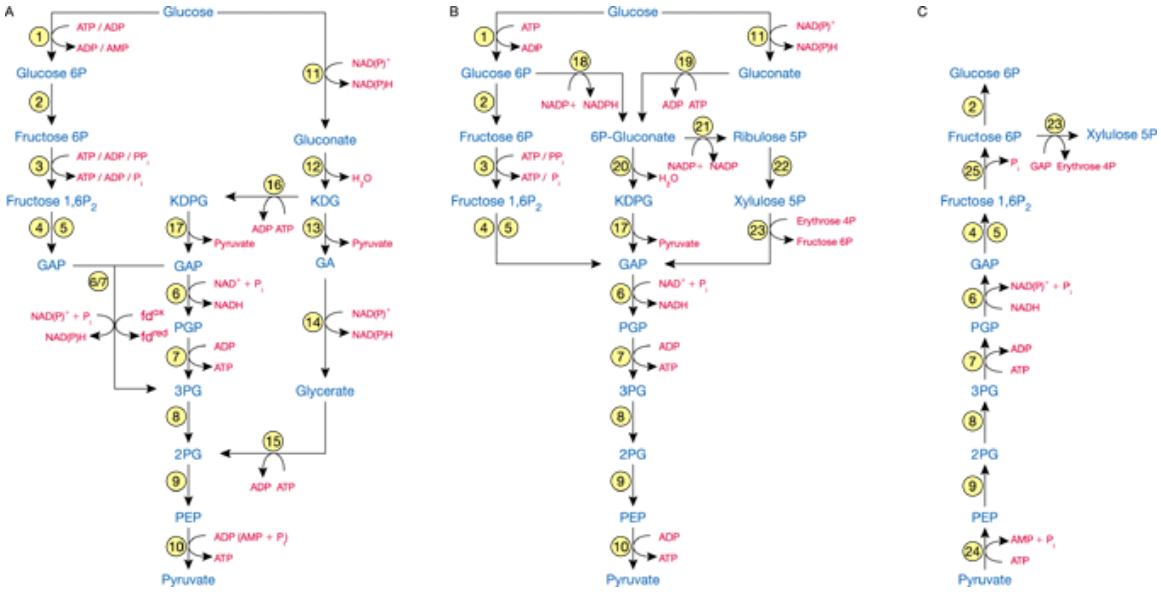

Figure 1.2 - (Verhees et al. 2003) (a) Glucose degradation in Archaea via variants of the EM and ED pathways. (b) Bacterial/eukaryal routes of glucose catabolism: EM, ED and oxidative PP pathway. (c) Gluconeogenesis in archaea and bacteria/eukarya results in the reduction of pyruvate to fructose 6-phosphate (Fructose 6P), which may be further converted into glucose 6-phosphate (Glucose 6P) or xylulose 5-phosphate (Xylulose 5P) via the non-oxidative PP pathway. Abbreviated metabolites in (a-c) are: GA, glyceraldehyde; Fructose 1,6P2, fructose 1,6-bisphosphate; PGP, 2,3-bisphosphoglycerate; 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; fdox, oxidized ferredoxin; fdred, reduced ferredoxin. Key to enzymes: (1) GLK; (2) PGI; (3) PFK; (4) FBA; (5) TIM; (6) GAPDH; (7) PGK; (6/7) GAPOR and GAPN; (8) PGM; (9) enolase; (10) PYK; (11) Glc DH/Glc-lac; glucose dehydrogenase/gluconolactonase; (12) G-hydr, gluconate dehydratase; (13) KDG ald, KDG aldolase; (14) Ald DH, glyceraldehyde dehydrogenase; (15) Gly kin, glycerate kinase; (16) KDG kin, KDG kinase; (17) KDPG ald, KDPG aldolase; (18) G6P DH/6PG-lac; glucose-6-phosphate dehydrogenase/6-phosphogluconolactonase; (19) GlcA-kin, gluconate kinase; (20) PG-hydr, 6-phosphogluconate kinase; (21) PG DHDC, 6-phosphogluconate dehydrogenase (decarboxylase); (22) RP epi, ribulose-5-phosphate 3-epimerase; (23) Tr.ketolase, transketolase; (24) PPS; (25) FBP.

Described metabolic pathways in archaea are unique in many ways compared to the classical pathways in bacteria and eukaryotes (Snijders et al. 2006). In *S. solfataricus*, glucose degradation is carried out in a non-phosphorylated version of the ED in which glucose is converted to pyruvate glucose dehydrogenase, gluconate dehydrogenase, gluconate dehydratase, KDG, glyceraldehydes dehydrogenase, glycerate kinase, endolase, and pyruvate kinase. Recent evidence has pointed towards the operation of a semiphospyorylated ED pathway in *S. solfataricus* in which KDG is phosphorylated (Ahmed et al. 2005).

The elucidation of unique metabolic pathways in archaea has yielded the characterization of new enzymes, and will most likely continue to do so. These enzymes, key in the breakdown and formation of numerous starches and sugars, hold great possibility for application in industrial processes that require increased temperature stability and/or adverse hydrogen ion concentrations. Extremophiles in general have been attractive to the Biotechnical industry as they hold the promise of a plethora of thermostable enzymes useful in unlimited applications.

Biotechnical Applications of Extremophiles

There are currently more than 23 completed genome sequences for Archaeal extremophiles (TIGR-CMR http://cmr.tigr.org/tigr-scripts/CMR/CmrHomePage.cgi) (Laksanalamai and Robb 2004). Although annotation of these genomes includes many unknown genes, the data have fueled much speculation about archaeal gene function, structure, and expression. The fact that extremophiles are capable of thriving under non-standard conditions led to the assumption that their enzymes have been optimized to

function in extreme environments. The data collected on many of the enzymes that have been isolated and functionally characterized from many extremophiles support this hypothesis (Madigan and Marrs 1997).

The biotechnology industry has shown great interest in extremophiles because they are possible sources of valuable commercial reagents such as enzymes that operate well out of the ranges of temperature, pH, or salinity that can be obtained with enzymes isolated from mesophilic organisms. Standard enzymes often stop working when exposed to harsh reaction conditions in industrial processing, so extra steps must be taken to achieve the desired results. An enzyme that operates outside the "normal" reaction range has the potential for increasing process efficiency by eliminating costly additional steps.

The most famous example of this is the discovery and isolation of *Thermus aquaticus* from Mushroom Springs (YNP) in 1966, which led to the discovery of Taq Polymerase. Taq Polymerase is a thermo-stable enzyme used in the polymerase chain reaction (PCR) that is able to replicate DNA at the elevated temperatures required for the process. This biotechnical innovation is arguably the largest contributing factor to the rapid advances made in molecular biology in the 1980s thereby jumpstarting the fledgling biotech industry. Processes such as the classification of 16S RNA led to the delineation of Archaea, the domain that has been that pivotal in creating the impetus for the topic of this study. In 1997 alone industry worldwide spent $2.5 billion on research and development of new enzymes (Madigan and Marrs 1997). These enzymes range from proteases developed from thermophiles and used in detergents to hydrolases used in

brewing and baking processes, to dehydrogenases developed from halophiles that are used as biocatalysts in organic media (Madigan and Marrs 1997).

Another useful application of knowledge that could be gained from the in-depth study of hyperthermophiles would be a better understanding of the mechanisms that are responsible for the increased stability required in the proteins of extremophiles. The question is, how do molecules in heat-loving microbes remain functional under conditions that destroy their mesophilic homologs? So far there is no one satisfactory answer for this question, though we do know proteins found in heat tolerant organisms tend to be posses more ionic and hydrogen bonds as well as more hydrophobic residues in their interiors than are found in the proteins of their homologs in mesophilic organisms.

Viruses

Viruses make up in shear mass the largest component of the biosphere and often outnumber their hosts in marine environments by an order of magnitude (Suttle 2005; Wommack and Colwell 2000). It is known that all cellular organisms are susceptible to infection, often by more than one type of virus, which would infer that viruses are the most diverse entities on this planet (Fuhrman 1999).

Basic research on viruses has been pivotal in the advancement of the field of biochemistry, cell biology, and molecular biology as well as a greater understanding of the fundamental processes of life as we know it. The most obvious examples of this are the discovery that DNA is the fundamental carrier of genetic information (Hershey and Chase 1952), and that mRNA is the intermediate molecule for the transfer of information

to the ribosome (Brenner 1961). Other contributions have been the discovery and mapping of the first gene (Benzer 1955); discovery of the mechanics of discontinuous DNA replication (Okazaki et al. 1968); and the discovery of restriction endonucleases (Nathans and Smith 1975). The value of these discoveries is exemplified by the direct link between the development of basic virology and the prosperity of the multibillion-dollar biotechnology industry (Wommack and Colwell 2000).

Viruses are found associated with every organism, in all three domains of life, in every habitat in nature investigated to date (Ackermann 1987, 2001). New virus descriptions usually include an electron micrograph, and surveys of these descriptions are the historical basis for virus taxonomy (Ackermann 2001). These surveys also illustrate the diversity of viruses, and open the door for speculation about the course of their evolution. The first comprehensive listing of viruses was published in 1967 and listed 111 negatively stained viruses—99 of which were tailed, nine cubic, and three filamentous (To et al. 1966). A survey completed in1995 totaled 4,389 tailed phages and 162 cubic, filamentous, and pleomorphic phages (Ackermann 1996). Approximately 500 more were added to the list by 2000, and by 2003 an estimated 5,300 bacterial viruses had been identified via electron microscopy (Ackermann 2003), making them the largest identified viral group in nature (Van Regenmortel et al. 2000).

The International Committee on Taxonomy of Viruses (ICTV) currently classifies viruses into three orders, and 78 families (ICTV 2006, http://phene.cpmc.columbia.edu/ICTVdB/index.htm). Bacteriophages themselves constitute one order, 13 families, and 30 genera. Viral morphologies include tailed,

polyhedral, filamentous, or pleomorphic and the largest percentage of phages have double stranded DNA (dsDNA) genomes. There are, however, smaller groups that are classified by their single stranded DNA (ssDNA), single stranded RNA (ssRNA), and double stranded RNA (dsRNA) genomes. A small percentage of these viruses contain lipid envelopes, either external or internal. At the latest count there were 4,950 viruses of the tailed variety that fall into the order of *Caudovirales*, and three large phylogenetically related families (Ackermann 2003). The balance of 190 filamentous, polyhedral, and pleomorphic viruses are classified into 10 small families that differ only by the most basic properties.

Viral families are largely defined by morphology, the type of nucleic acid they contain, and their general replication strategy. Because of the difficult nature of viral taxonomy (highly plastic genomes and variable morphologies), the ICTV uses every available property for classification and has adopted the "polythetic species concept." This means that a species is defined by a set of properties, some of which may be absent in an individual but apply to most of the members in general (Van Regenmortel 1990).

Evolutionary Significance

Viruses have been classified in more than 140 bacterial genera and they occur in Cyanobacteria, exospore and endospore formers, mycoplasmas, chlamydias, aerobes, anaerobes, budding, gliding, ramified, sheathed, and stalked bacteria (Ackermann 2003). Viruses occur in bacteria of all sections of the microbial worlds, however their arrangement according to the phylogenetic grouping of bacteria reveals that most major phyla are lacking defined phages. This is likely due to the fact that most phages have

been discovered in medically and industrially relevant bacteria that are easily grown and have been investigated extensively. This suggests that many more phages will be identified in the future as other bacteria are investigated.

If one considers that prokaryotic viruses outnumber their hosts in aquatic environments by an order of magnitude (Wommack and Colwell 2000), reaching levels as high as one hundred million per milliliter of sea water in the tropic zone, this leads to an awe inspiring presence in the biosphere as a whole. In the ocean alone it is estimated that there are ~4 x 10^{30} viruses containing 200 Mt of carbon (Suttle 2005). If these viruses were stacked end-to-end they would span ~10 million light years, which is ~100 times the distance across our galaxy (Suttle 2005). If this overwhelming viral presence can be assumed for all organisms, it means that all of life is completely immersed in a sea of viruses (Bamford 2003).

Viruses ubiquitous distribution suggests that they are ancient and may have played a central role in early cellular evolution (Forterre 2003). Genomic studies indicate that lateral gene transfer and non-orthologous gene replacement have been major forces in the evolution of cellular life, and this conclusion has led many to believe viruses may have been the vehicles for such transfers. This would elevate their importance in the greater scheme of the biosphere far above that of merely bearers of pathogenicity islands (Forterre 2003).

Not only do viruses and virus-like elements overwhelmingly out number cellular life forms, but also they are present in a significant portion of their genomes. Sequencing has revealed that cryptic and temperate phages are incorporated into the genomes most

prokaryotes, and in the portion of eukaryotic genomes that don't code for proteins, virus-like elements and their remnants are overwhelmingly present (Filee et al. 2003). Therefore, the ability a virus possesses to cross cellular walls and membranes designed to keep the rest of the world out, and then commandeer the cellular productive machinery to its own ends is, in and of itself one of the largest selective pressures that organismal life encounters.

In spite of their overwhelming presence in the biosphere, viruses remain largely understudied compared to their cellular hosts (Filee et al. 2003). The probable reason for this lack of study is the popular misconception that viruses are derived from cellular nucleic acid, and that they only recently became autonomous entities. Another factor may be the difficulty one encounters following a phylogenetic signal at the nucleic acid level due to the plastic nature of viral genomes. One extreme example of this is the rapidly changing genomes of viruses as studied over a time course in a single YNP thermal pool (Wiedenheft et al. 2004). Large portions of their genomes varied considerably over short periods of time (weeks). This is apparently an exaggerated feature of thermal archaeal viruses in comparison to the rest of the viral world (D. Prangishvili 2003), and may be due in large part to mixing of transient viral populations in these environments (Snyder et al. 2004). When long evolutionary time spans and particularly transient genomes are considered, it becomes necessary to look at the conservation of structure and function in order to rationalize the organization of viral life forms (Bamford 2003). Structure and function may be conserved over periods of time in which comparative viral nucleic acid sequences evolve to the point of no detectable

similarity above random genomic noise. Unfortunately this loss of signal at the genomic level cannot be easily remedied because of our current inability to translate amino acid sequences into primary and secondary protein structures, let alone protein-protein interactions.

The conservation of protein function and structure may be key in establishing relatedness between life forms in deep evolutionary time frames (Bamford 2003). This becomes especially relevant when considering the widely held belief that some form of molecular organization, i.e. folding of macromolecules, began before the solidification of what we now consider the earliest life forms (Nisbet and Sleep 2001). It is possible that viruses diverged as autonomous nucleic acid replication entities before the delineation of the three current domains of life (Filee et al. 2003). One argument that supports this view is the existence of a number of striking structural and functional similarities between viruses that infect different organisms belonging to separate domains of life. For instance, many tailed phages from both Archaea and Bacteria have striking morphological resemblances (Zillig et al. 1996). The archaeal SIRV has structural similarities to and shares functional replication mechanisms with eukaryal Poxviruses (Peng et al. 2001).

Crenarchaeal dsDNA Viruses

The morphological and genomic diversity found in studies of viruses that infect hyperthermophilic crenarchaea is unprecedented in the field of virology. To date, 21 different virus have isolated from, and propagated in, members of the genera *Sulfolobus, Acidianus, Thermoproteus,* and *Pyrobaculum* (Kessler et al. 2004; Ortmann et al. 2006).

These viruses have been assigned to five new families due to their unique nature: Spindle-shaped *Fuselloviridae*; flexible filamentous *Lipothrixviridae*; stiff rod-shaped *Rudiviridae*; droplet-shaped *Guttaviridae* (D Prangishvili et al. 2001); spherical *Globuloviridae* (Haring et al. 2004), and two proposed families—*Bicaudaviridae* and *Ampullaviridae* (Kessler et al. 2004; Prangishvili and Garrett 2005). These viruses are not only morphologically unique but also notable as a group in the sense that they have been shown to establish a productive infection of the host without killing or lysing it. There is one known exception to this in ATV, which is released upon lysis of the host (Haring et al. 2005). This carrier state differs from the majority of dsDNA viruses of bacteria and euryarchaea that eventually kill the host cell in the course of infection and replication.

SSV

The first archaeal virus to be described *Sulfolobus spindle-shaped virus* 1 (SSV1) is the founding member of the *Fuselloviridae* family. SSV1 is approximately 60 nm X 90 nm in dimension (figure 1.3) and was originally isolated in Beppu, Japan, from *Sulfolobus shibatae* (Yeats et al. 1982), and also infects *S. solfataricus* (Schleper et al. 1992). SSV1's genome is circular dsDNA in nature and 15.5 kb in length, and possesses 34 ORFs (figure 1.3). The viral genome has a low G+C content (39.7%), which is similar to that of its host. The genome is maintained in infected hosts both as negatively and positively supercoiled in episomal form, and as an integrated provirus (Schleper et al. 1992). The integrated viral DNA may serve as a template for transcription, while the

negatively supercoiled form is used during replication and the positively supercoiled form for encapsidation (Snyder et al. 2003). This was the first virus to be described that contained DNA in a positively supercoiled form (Nadal et al. 1986). SSV1's genome integrates into an arginyl-tRNA gene of the host (Reiter and Palm 1990). *Sulfolobus* can be transfected with SSV1 DNA (Schleper 1993), and the virus has been used to construct several shuttle vectors (Stedman et al. 1999a).

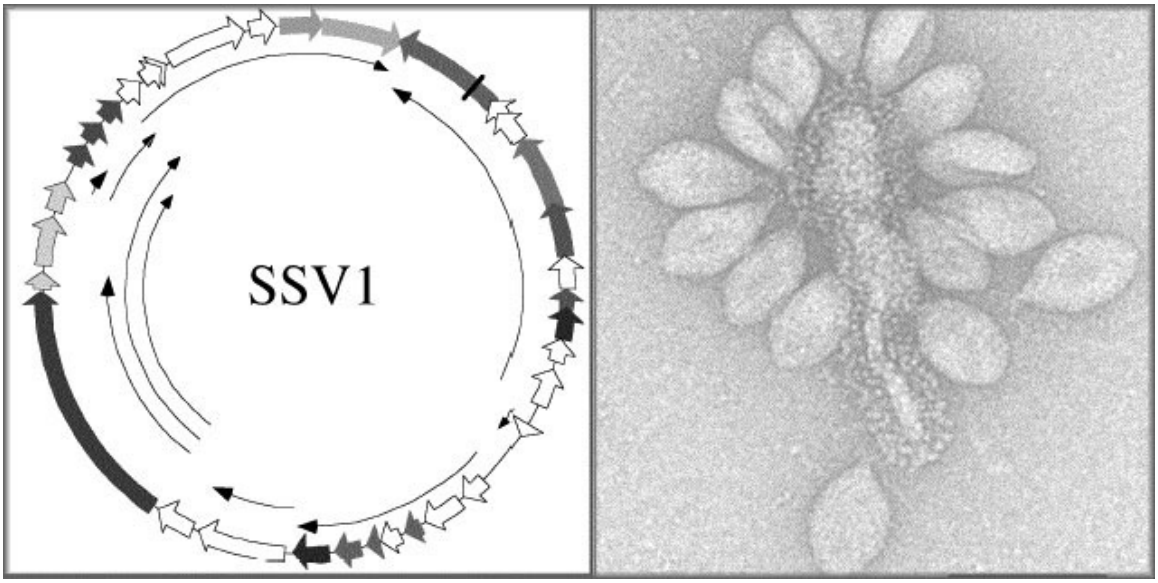

Figure 1. 3 - Genome map of SSV1 showing 34 ORFs, and TEM of the virus (Palm et al. 1991).

There are only four genes in SSV1 that have been identified with function. VP1 and VP3 are viral coat proteins; VP2 is a DNA binding protein, and another ORF that is the viral integrase. None of the other ORFs match genes in the public databases, but they are currently being investigated by structural analysis using X-ray crystallography. Nine

transcripts cover all 34 ORFs suggesting that the viral genes are transcribed in a polycistronic manner (Reiter et al. 1987a).

A second SSV (SSV2) was isolated from a solfataric field in Reykjanes Iceland and infects *Sulfolobus icelandicus* (Arnold et al. 1999a). This virus is nearly identical to SSV1 in morphology and size, but its genome is slightly smaller at 14.8 kb. The genome is 55 percent similar to SSV1 and the ORFs are arranged in a similar order (Stedman et al. 2003). VP1 and VP3 are conserved between SSV1 and SSV2, however VP2 (the DNA binding protein) is lacking in SSV2. The SSV2 genome integrates into a glycyl-tRNA gene instead of an arginyl-tRNA gene (Wiedenheft et al. 2004).

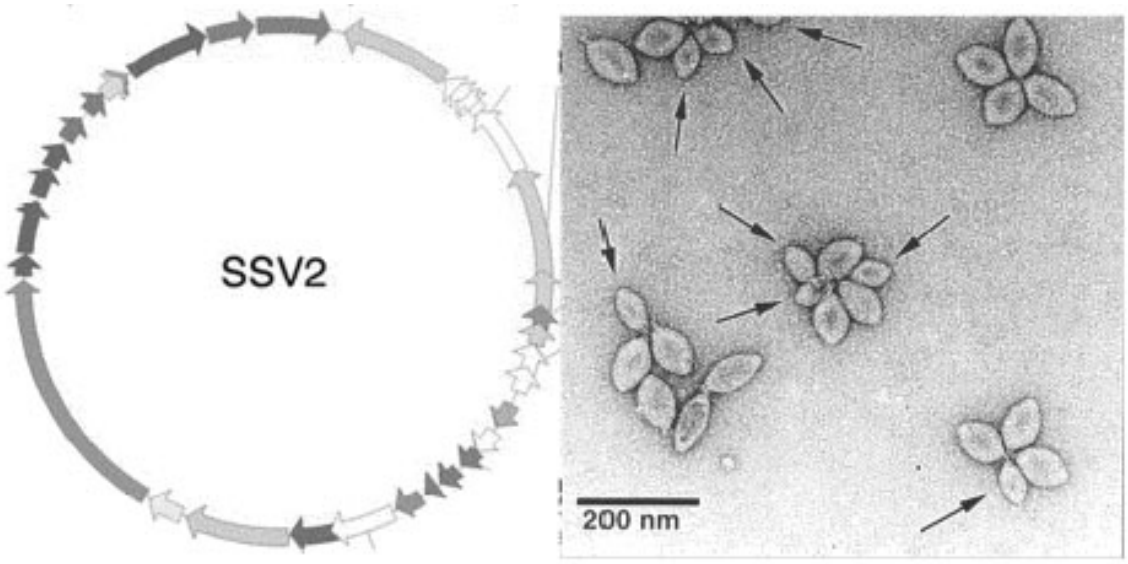

Figure 1.4 - Genome map of SSV2 and a TEM of the particle (Stedman et al. 2003).

Two other SSV types have isolated and sequenced, and a genome comparison has been performed among all of the SSVs. One of these is from a spring in the Ragged Hills area of YNP and the other from Kamchatka, Russia (Wiedenheft et al. 2004). This comparison shows approximately one half of the genome on all four viruses to be similar

to one another and the other half of their genomes show no similarity to other sequences in the public databases.

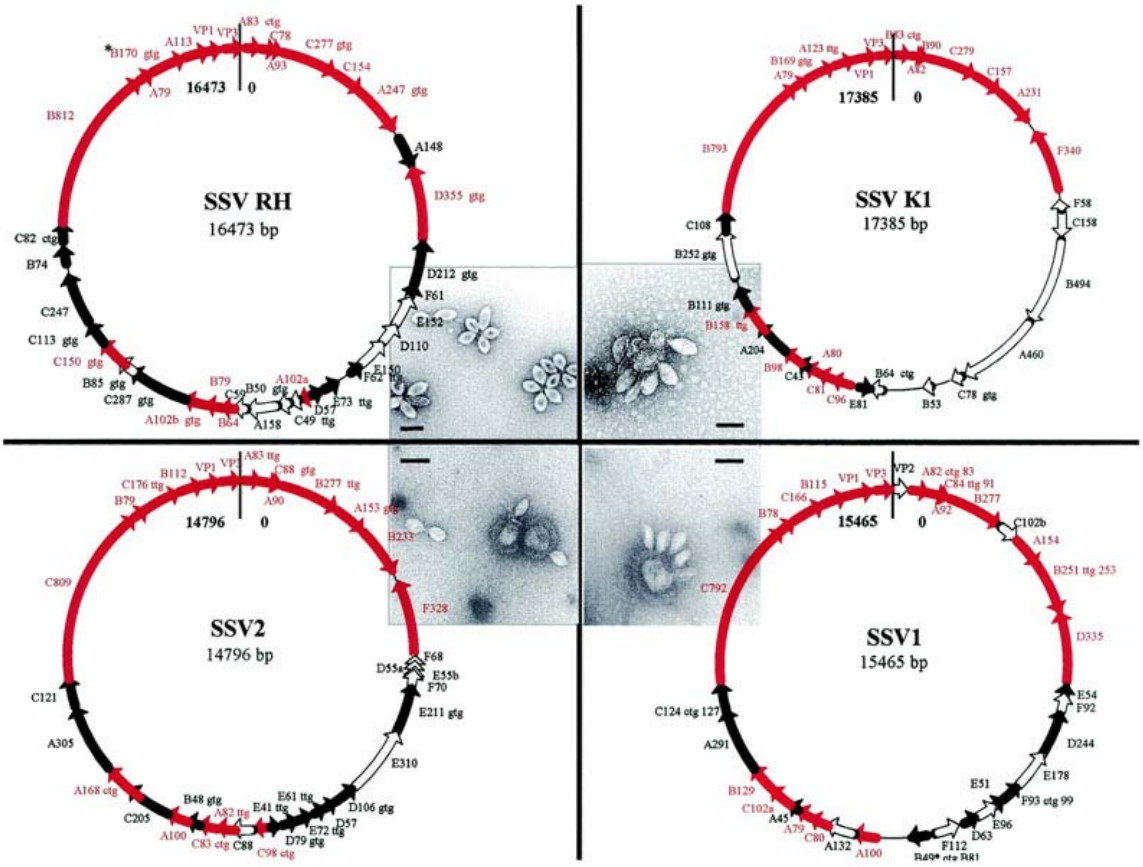

Figure 1.5 - (Wiedenheft et al. 2004) Genome maps of the four isolates. Conserved ORFs shared by all genomes are shown as red arrows. ORFs shared between two or three of the SSV genomes are shown as solid black arrows, and ORFs unique to each isolate are shown as open arrows. TEM images of each virus are positioned next to maps of their respective genome (bars, 100 nm). The alternative initiation codons (asterisks) are indicated directly following the name of each ORF in which they were identified.

SIRV

The two rudiviruses, *S. icelandicus rod-shaped virus*, SIRV1 and SIRV2 are rod-shaped with dimensions of approximately 25 nm by 400 nm (figure 1.6). The linear

dsDNA genome of SIRV1 was sequenced and determined to be 32,301 bp in length while SIRV2's genome contains 35,502 bp. At both ends of a unique core sequence in the viruses there is a flanking sequence of 2,029 bp and 1,628 bp respectively that are identical to each other and form perfect inverted terminal repeats (ITRs). Both ITR's are separated from the core sequence by 70 bp segments that are also very similar to one another in sequence. The sequence at the ends of the linear DNA is uninterrupted suggesting that the double strands are connected at these points and fold back upon themselves. On either side of the point where the DNA strand doubles back and transitions into the complimentary strand the sequences are completely complementary to one another (Blum et al. 2001). Analysis of another virus that also infects *Sulfolobus, S. icelandicus filamentous virus* SIFV (described below), shows that portions of its sequence are similar to SIRV and it also contains long ITRs at the ends of its genome.

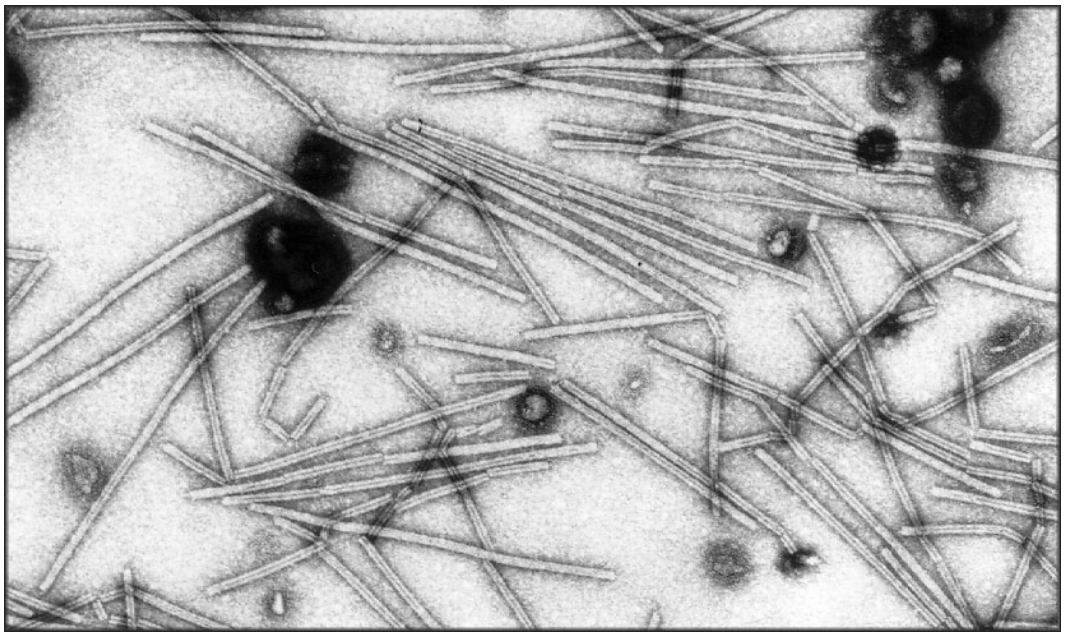

Figure 1.6 – TEM of SIRV isolated from Yellowstone National Park.

Both SIRVs have low GC content genomes (commonly denoting a low melting temperature of double stranded nucleic acid) at 25%. This is even lower than that of other *Sulfolobus* viruses such as SIFV at 33% (Arnold et al. 2000b) and SSVI at 40% (Palm et al. 1991).

The SIRV viruses have similarities to the eukaryal poxviruses (Baroudy and Moss 1982), African swine fever virus, and *Chlorella* viruses (Zhang et al. 1994). The eukaryal viruses differ slightly from SIRV in that their ends are not part of the ITRs and not completely base paired. Regions forming hairpin loops exist at each end of the genomes of these viruses and are reverse and complementary in sequence to one another.

The genomes of both SIRV1 and SIRV2 are stable when replication occurs within the original host. However, when SIRV1 is passaged into a new host it produces variant genomes (Prangishvili et al. 1999b) while SIRV2 remains stable under the same conditions. After multiple passages through the new host, unstable variants became genetically stable. It was initially estimated that the mutational rate in the genome of SIRV1 was 10^{-3} substitutions per nucleotide per replication cycle (Prangishvili et al. 1999b), which is abnormally high for a DNA virus.

The working hypothesis for SIRV1s genomic instability proposes that genomic variability probably is not a result of high mutational rates, but instead arises from selection for different genomes. Therefore, the abnormally high genomic variation observed in this virus is due at least in part to the existence of and co-replication of numerous viral genomes in the host cell (Peng et al. 2004). In other words, SIRV1 exists as a mixture of variants with one or more of them dominating the population in a given

Sulfolobus host. This genomic variability would give the virus the ability to rapidly adapt to new hosts.

SIRV1s genomic adaptability is initiated when the virus encounters changes in its environment, such as a new host, and is facilitated by a novel intron-driven mechanism that creates genomic variability and recombination between differing genomes (Peng et al. 2004). Preponderance of the genomic variability in this virus leads to larger questions involving viruses in hot environments in general—are multiple infection events per cell common, and if they are, is this a prerequisite for the formation of numerous variant genomes? Further investigation of these questions may shed light on the nature of the large genomic variability observed in SSV-like viruses by Snyder et al. over time in a single thermal pool (Snyder et al. 2004). One wonders if this observed variability accounts for the dissimilarity of genomes in viruses of hot environments in general, or if it is a function of geographic isolation of these thermal environments. More than 90% of all ORFs of sequenced thermophilic viral genomes have not been assigned functions and show no significant homology to sequences in the public databases (Kessler et al. 2004).

Many studies have observed the controls of viral gene expression in hyperthermophilic archaea but the mechanisms therein have not been elucidated (Prangishvili et al. 2001). The first systematic study of transcription in thermal viruses over the replication cycle was recently carried out on SIRV1 and SIRV2. The identification of promoter sequences in their viral genomes confirms the original observation of the similarity between TATA box-containing promoters of the eukaryal

RNA polymerase II and the Archaeal transcription machinery (Reiter et al. 1987b; Reiter et al. 1988).

Highlights of the SIRV study include the finding that the time required for a 50% infection rate of the host cells at a multiplicity of infection (MOI) of five was 14 minutes for both viruses, and it took four hours and 2.5 hours respectively for the appearance of intracellular virus particles in SIRV1 and SIRV2. The period between infection and the first release of the virus from the host was eight hours for SIRV1 and six hours for SIRV2. It was shown that transcription started from 21 sites in SIRV1 and 29 sites in SIRV2 at nearly the same time, and transcripts came from both strands of the genomes about equally. The length of the transcripts is indicative of clustering of viral ORFs into operons with the ensuing transcription of polycistronic mRNA. About 50% of the transcription start sites were positively identified with TATA boxes along with a purine-rich region corresponding to the transcription factor B binding site that is indicative of *Sulfolobus* promoters.

Similar genome organization/architecture and replication strategies make a plausible argument for a phylogenetic relationship between the rudivirus family and the aforementioned eukaryal viruses. Furthermore some bacterial and eukaryal viruses are similar in their replication processes and genome organization (Forterre et al. 1992). Similarities in capsid protein structures have also provided evidence for links between eukaryal and bacterial viruses (Hendrix 1999), and now also to an archaeal virus (this work).

It has been previously argued that it is unlikely that commonalities between viruses from separate domains of life can be explained by the transfer viruses between the members of these different domains (Prangishvili et al. 2001; Zillig et al. 1998). This line of reasoning holds that it would require an unlikely number of biochemical adaptive changes in order for viruses to jump from one domain to another. These barriers include the differences in a host's defenses required for survival in radically different environments, as well as the incompatibility of transcription mechanisms. It can be argued that it is possible that the ancestors of these viruses, or recognizable parts therein, existed before the divergence of the three domains of life Zillig et al. 1998; Zillig et al. 1996).

Other Thermal Viruses

Pyrobaculum spherical virus (PSV) is the first virus to be described for the *Pyrobaculum* genus. *Pyrobaculum* is another thermophilic archaeon in the crenarchaeal order Thermoproteales. In investigating the host range of PSV it was found that it would also propagate in *Thermoproteus tenax* (Haring et al. 2004), which is not the closest neighbor in a neighbor-joining analysis of 16S rDNA. Morphologically, PSV was found to have a spherical diameter of 100nm, sporting variable numbers of spherical protrusions 15 nm in diameter with no discernable symmetry of placement on the virion. Long filaments, 6nm in diameter (figure 1.7d), were observed protruding from disrupted particles that have closed loops at the ends. Fourier analysis of the filaments indicated a helical superstructure, which suggests that the virions are enveloped and contain a tightly packed nucleoprotein in a super-helical conformation. If this structural nucleoprotein

architecture was confirmed, it would be the first example of its kind in an enveloped DNA virus.

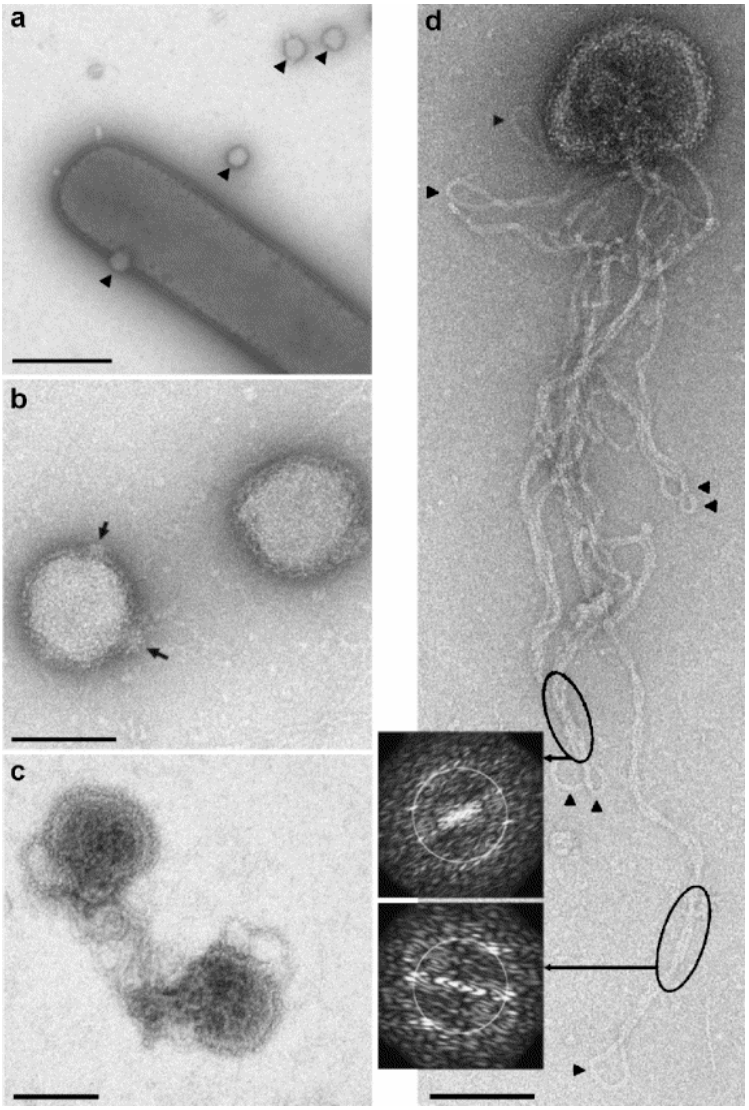

Figure 1.7 - (Haring et al. 2004) Electron microscopy of *Pyrobaculum* sp. D11 and PSV virions negatively stained with 3% uranyl acetate. (a) Portion of a *Pyrobaculum* sp. D11 cell with four PSV virions marked by arrowheads; scale bar: 0.5 mm. (b) Two intact PSV virions, spherical protrusions are marked by arrows; scale bar: 0.1 mm. (c) Two disrupted PSV virions extruding disordered filamentous material; scale bar: 0.1 mm. (d) Disrupted PSV virion with extended filaments extruding from the particle. Arrowheads mark several loops, and Fourier analysis analyzed two stretches enclosed by ellipsoids as shown in the insets. The circle indicates a frequency of (2.8 nm)-1; scale bar: 0.1 mm.

An interesting characteristic of PSV is that its genome has a concentration of genes on only of its strands, with only three residing on the opposing strand (Haring et al. 2004). In this way, PSV differs from most of the archaeal viruses with linear dsDNA genomes. PSVs genome does however possess ITR's at its ends similar to those of the rudiviruses, but its ORF's have almost no significant matches with anything in the database, including the other crenarchaeal viruses.

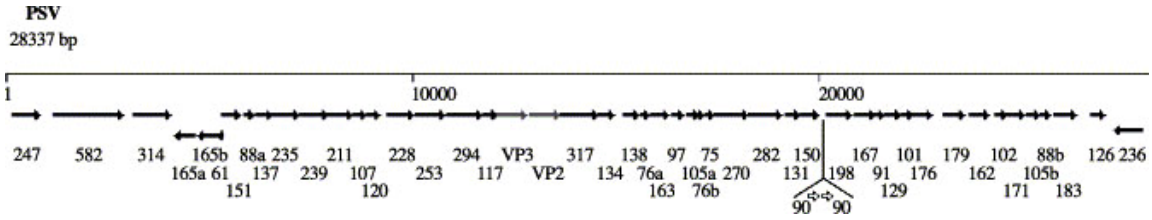

Figure 1.8 - (Haring et al. 2004) Phylogenetic tree for Crenarchaeota determined by a neighbor-joining analysis of 16S rDNA sequences showing the position of *Pyrobaculum* sp. D11, the host of PSV, in bold type. The scale bar represents 0.10 fixed mutations per nucleotide position.

Acidianus filamentous virus 1 (AFV1**)** is the first virus of *Acidianus* to be described in detail. Its host range is confined to several strains of *Acidianus,* and the filamentous virus measures 900 nm X 24 nm. It possesses claw-like structures at both ends that attach to the host pili and measure ends that are 20 nm in diameter. It has been assigned to the family Lipothrixviridae based on morphologic similarities. Like many of the other viruses of Crenarchaeaota it doesn't appear to lyse its host during a productive infection.

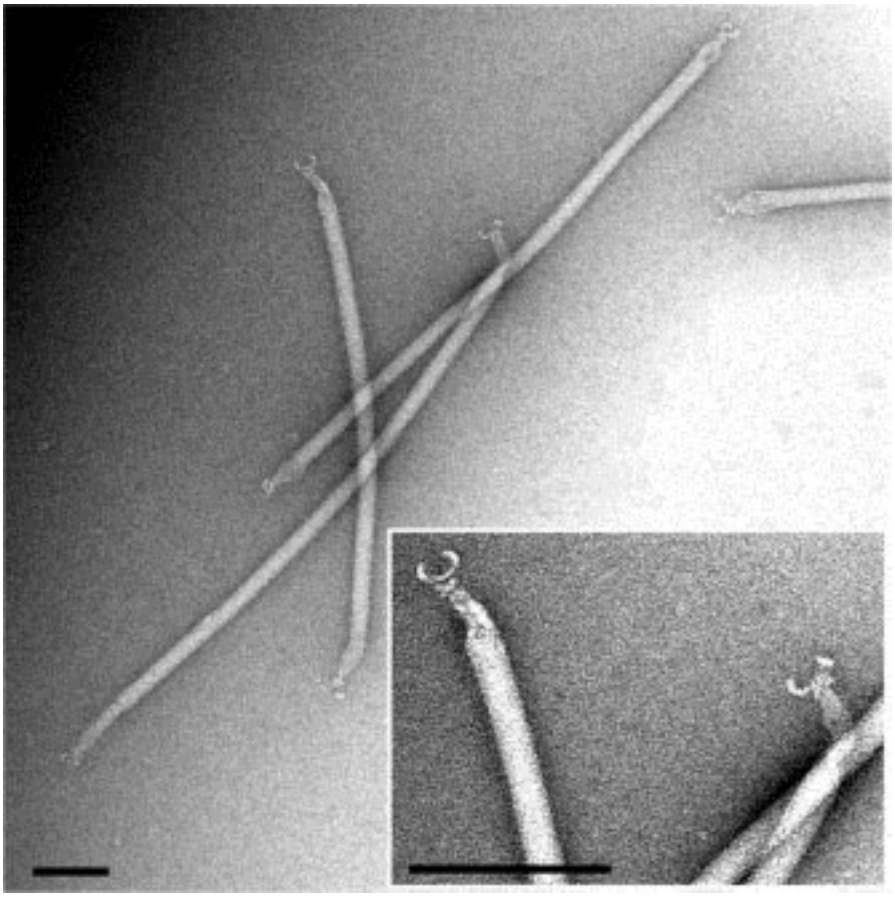

Figure 1.9 - (Bettstetter et al. 2003) Electron micrographs of particles of AFV1 with tail structures in their native conformation, negatively stained with 3% uranyl acetate. Bars are 100 nm.

Fifteen of AFV1's ORFs have homologs in SIFV's genome, as well as a 15% homology to ORF's in SIRV1 and SIRV2. A few of the ORF's that have homologs in SIFV are predicted to be glycosyltransferases, which could be involved in modification of host lipids. The terminal sequence regions of the genome are unique in the sense that they show smaller and less regular ITRs and several closely repeated direct repeats of TTGTT over the terminal 300 bp. This is reminiscent of telomeric ends of eukaryal

chromosomes where the telomerase creates multiple imperfect repeat sequences while generating a 3’ overhang that prevents shortening of the chromosome during replication.

Rachel et al. report on two hot pools in YNP (one acidic and one neutral: Crater Hills and Obsidian Pool respectively), in a study performed after the initial viral survey for this body of work was completed. They report an unprecedented viral diversity in enrichment cultures, and give an account of nine different virus like particle (VLP) morphotypes from the enrichment of the acidic pool and five from the neutral pool (Rachel et al. 2002). In the culture established from the acidic sample they report three virus morphotypes similar to known hyperthermophilic Archaea. These are stiff rods measuring 1,030 nm x 23 nm like the SIRV viruses of *Sulfolobus*; 800-950 nm x 24 nm filamentous particles resembling TTV2 of *Thermoproteus*; and 80 nm x 60 nm spindle-shaped viruses like the SSV viruses of *Sulfolobus*. The six other morphotypes that were not previously observed in Archaea include a short, helical, rigid rod about 140 nm x 23 nm similar to the Tobacco Mosaic virus; a filamentous particle 700-850 nm x 23 nm with bulbous ends; other filamentous particles measuring 1550 nm x 23 nm with rounded tips; zipper-particles made up of triangular sub-units that measure 100-200 nm x 15 nm; and pleomorphic particles that have either a tapered bulbous end 180 nm x125 nm in dimension with a tail 24 nm wide and 100-620 nm long, or the same bulbous body with the appendages on both ends.

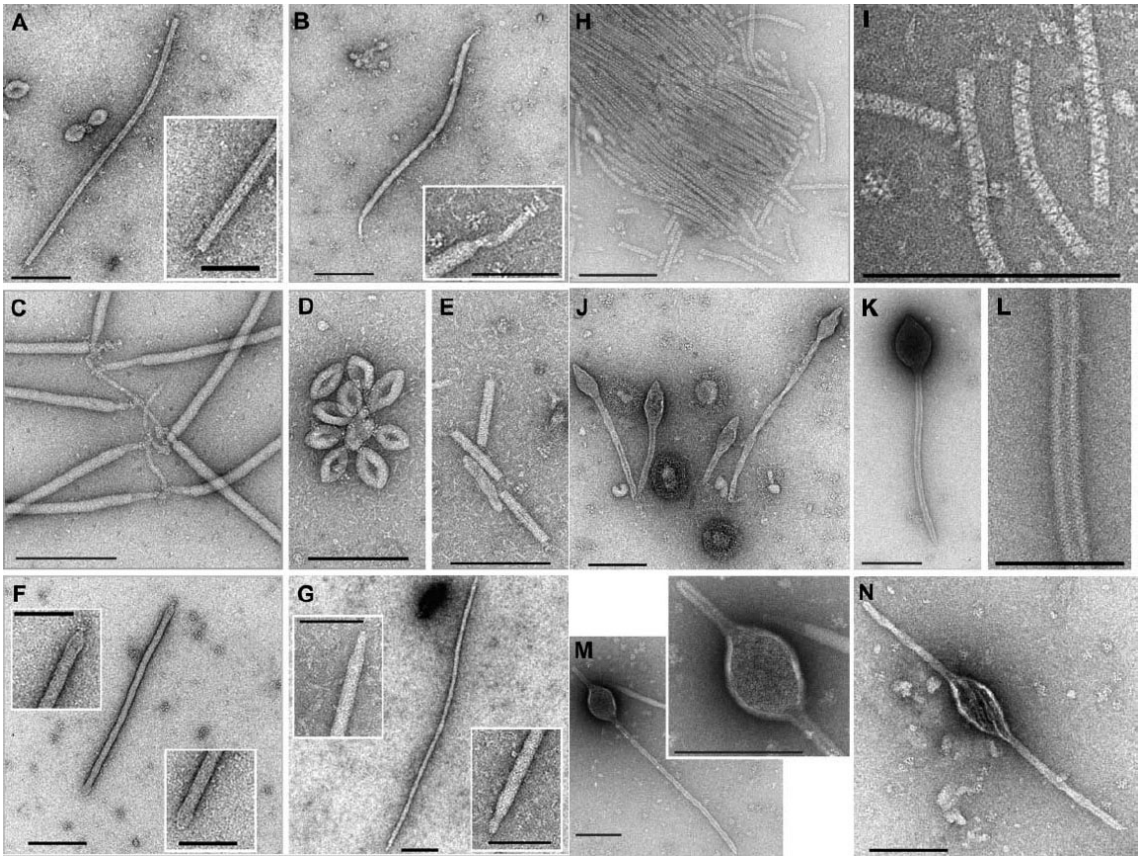

Figure 1.10 - (Rachel et al. 2002) Transmission electron micrographs of viruses and VLPs found in an enrichment culture of a sample from the Crater Hills region of YNP. (a) rod-shaped particle; inset: enlarged terminus; (b) filamentous particles with flexible, elongated terminal structures; inset: enlarged particle end; (c) often found attached to thin filaments; (d) spindle-shaped particles; (e) short rod-shaped particles; (f) filamentous particles with bulbous termini, both shown enlarged in the inserts; (g) long filamentous particles with rounded ends, both shown enlarged in the inserts; (h) aggregated zipper-shaped particles; (i) four zipper-shaped particles shown at high magnification; (j) pleomorphic particles with arrow-shaped heads and helical tails; (k-n) ellipsoid particles, (k) with one helical appendage, which is shown enlarged in (l), (m) with two helical appendages; (n) with twisted head and two helical tails. All bars are 200 nm except inserts, which are 100 nm.

In the enrichment from the neutral pool they report seeing typical head and tail morphologies that had not been previously detected in high temperature environments. As in the other sample, they saw both the stiff rod-shaped viruses like SIRV and filamentous particles similar to the filamentous virus of *Thermoproteus*, TTV2. They

also observed the odd shaped particles reported in the acidic enrichment with the tapered bulbous central feature and either one or two tails, but with differing dimensions of the tails and a distinct terminal structure.

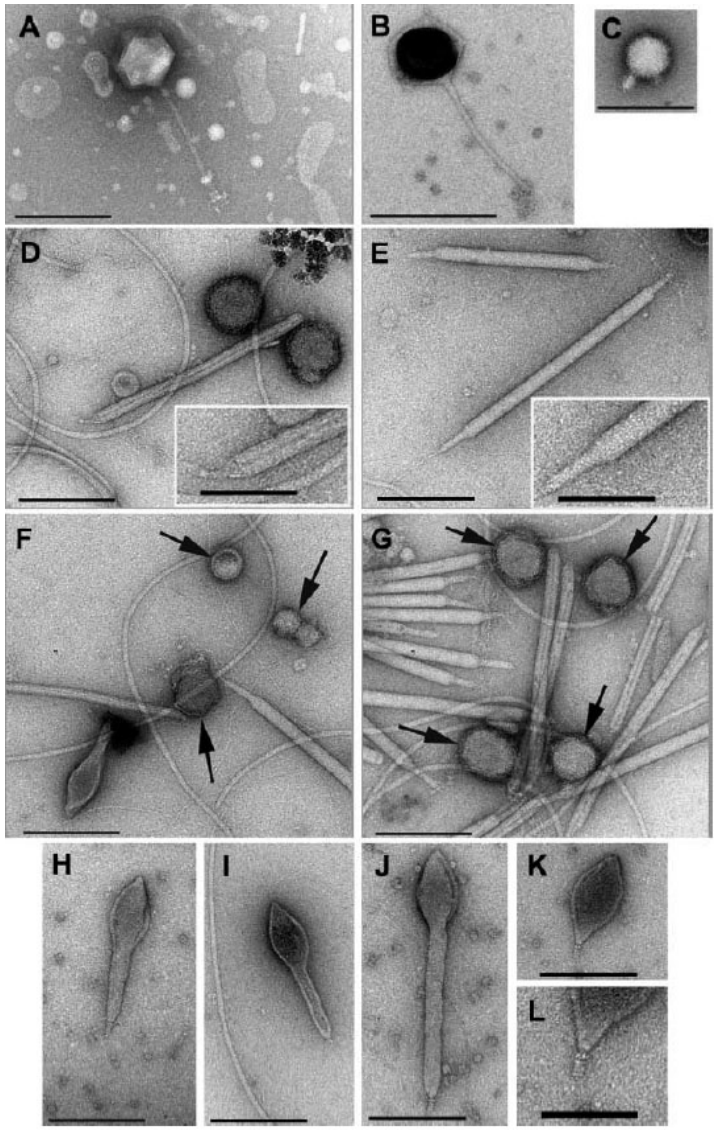

Figure1.11 - (Rachel et al. 2002) TEMs of viruses and VLPs found in an enrichment culture of a sample from Obsidian Pool in YNP. (a,b) head-and-tail viruses; (c) possible head-and-tail viruses with short tail; (d) rod-shaped particle; insert: enlarged particle end; (e) filamentous particles with elongated terminal structures, shown enlarged in the insert; (f,g) spherical particles with different diameters; (h-k) pleomorphic particles with arrow-shaped heads and wide tails with specific terminal structures enlarged in (l). All bars are 200 nm except inserts, which are 100 nm.

None of the reported viruses or VLPs above were shown to be harbored by *Sulfolobus*, although four of the morphotypes are similar to the characterized viruses of *Sulfolobus*. Hosts for at least three of the observed VLPs were shown to be strains of *Acidianus*.

Viral Morphologies and Diversity From Marine Environments

In another recent survey of thermal viruses, Geslin et al. report nine distinct virus morphotypes from enrichment cultures of samples taken from deep-sea hydrothermal vents in four different locations (Geslin et al. 2003). Eighty-nine enrichment cultures representing the geographically distributed hydrothermal sites yielded 15 virus-positive cultures, of which four contained multiple VLP morphologies.

Geslin et al. attempted to sort out the viral morphological soup by clustering them into two separate classifications. The first group consisted of stiff and filamentous particles of varying lengths, all of them resembling SIRV and SIFV to some degree. The possibility exists that the different lengths they report as evidence for the multiple morphotypes could be attributed to breakage of longer particles during sample preparation with the exception of one particle type that they stated could be a rigid rod with a club-like end. However, it's also possible that this is a variation of one of the particles they place in the second group.

The morphologies comprising the second group are all variations of the spindle shaped morphology. These VLPs vary in dimension from the standard SSV-like particle seen in typical rosettes, to the types reported above (and seen in numerous enrichments by this author) that have a larger than SSV-size spindle shape on one end with an

elongated tail, or in the middle with two tails. They report dimensions of the spindle with one tail at 100 nm x 50 nm and total particle length of up to 300 nm, and 100 nm x25 nm bulges with total lengths up to 430 nm for the double-tailers. Geslin et al. were unable to make a definitive determination of the host or hosts species of these VLPs, but based on observation they speculate that they belong to the order of *Thermococcales*.

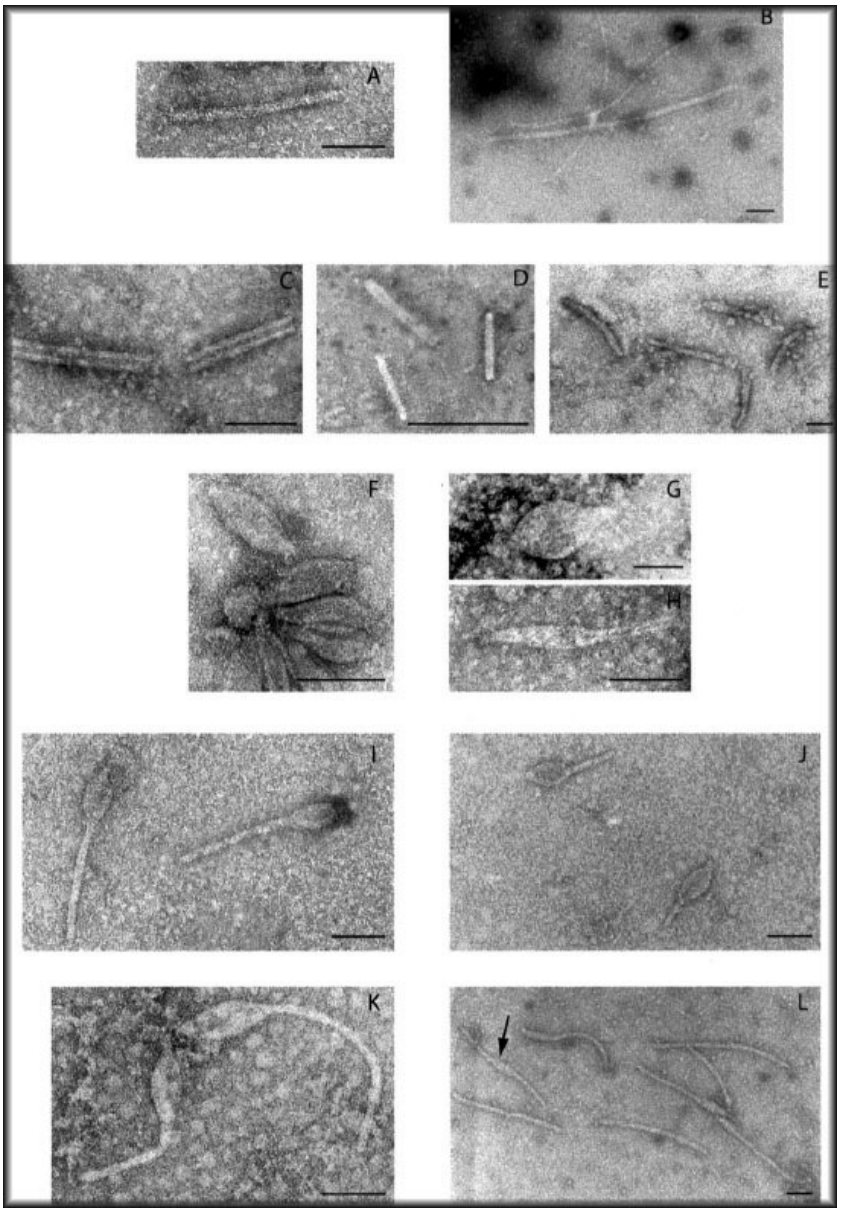

Figure 1.12 – VLPs from marine thermal environments. Measurement bars are 100 nm.

Table 1.1 - Modified from Snyder et al. (2003), viruses of the Crenarchaeaota.

Virus	*Host*	*Dimensions (nm)*	*Genome Size*	*References*
SSV1	*Sulfolobus shibatae*	spindle 60/90	ccc 15.5kb dsDNA	Martin 1984
SSV2	*S. islandicus*	spindle 55/80	ccc 14.8kb dsDNA	Arnold 1999 Stedman 2003
SSV3	*S. islandicus*	spindle 55/80	ccc 15kb dsDNA	Zillig 1998
SSV-RH	*S. solfataricus*	spindle 60/90	ccc 16kb dsDNA	Wiedenheft 2004
SSV-K1	*Sulfolobus* sp.	spindle 60/90	ccc 17.8 kb dsDNA	Wiedenheft 2004
SIRV1	*S. islandicus*	stiff rod 780/23	linear 32kb dsDNA	Prangishvili 1999
SIRV2	*S. islandicus*	stiff rod 900/23	linear 35kb dsDNA	Prangishvili 1999
SIRV-Y1	*S. solfataricus*	stiff rod 900/25	linear 33kb dsDNA	Rice 2001
SIFV	*S. islandicus*	filamentous 1950/24	linear 41kb dsDNA	Arnold 2000b
SNDV	*S. neozealandicus*	droplet 110-185/95-75	ccc 20kb dsDNA	Arnold 2000a
STIV	*Sulfolobus* sp.	icosahedral 70	ccc 19kb dsDNA	Rice 2004
TTV1	*Thermoproteus tenax*	flexible rod 40/400	linear 15.9 dsDNA	Janekovic 1983
TTV2	*T. tenax*	filamentous 20/1250	linear 16kb dsDNA	Janekovic 1983
TTV3	*T. tenax*	filamentous 30/2500	linear 27kb dsDNA	Janekovic 1983
TTV4	*T. tenax*	stiff rod 30/500	linear 17kb dsDNA	Zillig 1988
PSV	*Pyrobaculum* sp. *Thermoproteus* sp.	spherical 100nm	linear 28kb dsDNA	Haring 2004
DAFV	*Acidianus ambivalens*	filamentous 2200/27	linear 56kb dsDNA	Zillig 1994
AFV1	*Acidianus* sp.	filamentous 900/24	linear 21kb dsDNA	Bettstetter 2003

Viruses are the most common biological agents in the sea, typically numbering 10^9 per liter. Studies have shown that they are the most abundant biological entities in the sea at about 5-25 times the bacterial numbers in surface waters (Fuhrman 1999). Viruses follow the same abundance patterns as bacteria, decreasing by about an order of magnitude between rich costal waters and the nutrient poor open ocean. Abundance further decreases 5-10 fold from the euphotic zone to the upper mid-waters (500 m depth), and another several fold decrease to abyssal depths. Sea ice is highly enriched in viruses compared with the water beneath it, and sediments are highly enriched compared to the water directly above them (Paul et al. 1993).

It has been shown that viral numbers in marine environments are dynamic and correlate closely with ecological conditions such as algal blooms (Borsheim et al. 1990), providing evidence that they are active members of the microbial community. It is

thought that large fluctuations of virus numbers over short time periods (hours to days) is the result of large scale host cell lysis accompanied by rapid degradation of virus particles (Tuomi et al. 1997). Virus numbers have been tied to prokaryotic abundance (both bacteria and archaea) over hundreds of kilometers, indicating that most marine viruses infect these bacteria and archaea (Paul et al. 1993). In deep-sea environments, as much as half of the prokaryotic population may belong to the archaeal domain (Fuhrman 1999).

Summary

1. A survey of Archaeal viruses in the high temperature environments of Yellowstone National Park had not been conducted before this work.
2. Although viruses had been characterized from thermal areas around the world, relatively little is known about the overall diversity of these viruses.
3. Almost nothing is known about how the viruses and their hosts, residing in high temperature acidic environments, are able to thrive under such seemingly adverse conditions.
4. Very little is known about replication cycles of theses viruses and how they may interact with their hosts in extreme environments.
5. The objective of this study is to conduct the first comprehensive survey of viruses in the high temperature acidic environments of YNP, and to characterize a new virus with the aim of starting to answer some of the above questions.

CHAPTER 2

THE SEARCH FOR ARCHAEAL VIRUSES IN HIGH TEMPERATURE ACIDIC ENVIRONMENTS

Previous to studies described here, the systematic search, isolation and characterization of viruses in the high temperature environments found in Yellowstone National Park (YNP) had not been attempted. Previous studies carried out primarily by the Zillig lab reported the finding of viruses in thermal features such as mud pots, fumaroles, geysers, and hot pools in other parts of the world such as New Zealand, Iceland, Italy, and Japan (Arnold et al. 2000a; Zillig et al. 1994; Zillig et al. 1996).

Studying viruses in any given environment has traditionally been a productive scientific endeavor. It is through the study of viruses and virus-host interactions that some of the most fundamental biochemical underpinnings of life have been elucidated. These scientific milestones include the discovery of DNA, as well as many of the mechanisms responsible for the replication of this central molecule. It is by understanding the relatively simplistic nature of viruses in their interactions in a given environment that the workings of the more complex molecular machinery of their hosts can be pieced together. This is particularly intriguing in extreme thermal environments as the biochemical adaptations necessary for life to exist and thrive under such harsh conditions might hold clues to the stabilization of numerous bimolecular interactions outside of what one might view as normal reactive conditions.

With this in mind, we set out to determine if there were in fact viruses present in these thermal environments; and if so, what populations existed and what they might tell us about their residences.

Collecting Virus in Marine Environments

Estimation of viral abundance in marine environments has been facilitated by direct enumeration methods originating with the use of TEM. The first direct counts were reported by Torella et al. in 1979 when they used TEM to count viruses associated with plankton samples (Torrella and Morita 1979). Their original methods consisted of counting the viruses associated with plankton in the fraction of a sample that was retained after filtering with a 0.2 μm pore-size filter. The viral abundance estimates obtained in this manner were in the vicinity of 10^4/ml. In 1988 it was shown with epifluorence microscopy in conjunction with dsDNA binding fluorochromes (such as DAPI) that these numbers were more likely closer to 10^6/ml (Wommack and Colwell 2000).

The most commonly used method for counting virus particles with the TEM is the direct sedimentation of particles from unfiltered glutaraldehyde-fixed water samples onto fine-meshed Formvar-coated copper grids. Pre-filtration of water samples is not generally employed unless necessitated by high levels of suspended particulate matter. When samples are filtered, adjustments must be made to estimate the amount of virus that may have been lost due to adhesion to the filtered particulate and the filter itself (Wommack and Colwell 2000). There are other technical issues to be considered when pelleting particles onto an EM grid with a centrifuge, such as the fact that viruses do not sediment in parallel paths and therefore tend to concentrate toward the bottom edges of

the centrifuge tube. Also, different viruses have sedimentation rates depending on the overall shape and size of the particles, so it is difficult to be sure that all particles in a solution have been deposited on the grid. Due to the high magnification required to identify virus particles on TEM grids, the practical lower limit for particle concentration is 10^5 per ml. This limitation precludes the method from being useful in oligotrophic environments without modifications such as pre-concentration of the samples (Wommack and Colwell 2000). It is also estimated that a count variation of 20-25% is suffered due to uneven staining and variable distribution of the viruses on the grid.

In environments where low virus numbers are a problem, samples can be concentrated by ultra-filtration. Methods for concentration of viruses by ultra-filtration that have been developed include vacuum filtration (Wommack et al. 1996); centrifugal filtration (Fuhrman et al. 1993); pressure filtration through either a hollow-fiber filter or a spiral wound membrane filter (Chen et al. 1996; Suttle et al. 1991); and vortex flow filtration (Paul et al. 1991), a variant of pressure filtration. All of these methods have been shown to introduce a selective bias while improving detection limits. Overall, virus number estimates using unprocessed samples have been as much as three times higher than those using concentrated samples (Wommack and Colwell 2000). However, allowing for the fact that concentration may skew counts, the method has allowed for the investigation of viruses in environments where numbers are low.

Researchers studying marine viruses have shown that it is possible to do viral counts using epifluorescence microscopy with dyes that stain nucleic acid such as DAPI, YoPro, and Syber Green (Noble and Fuhrman 2000). This method is rapid and

inexpensive compared to performing virus counts with TEM. Viruses stained in this manner can be counted in the field within hours of collection using epifluorescence microscopy. Counts obtained in this manner have been shown to be similar to or 1-1.5 times higher than those with TEM (Noble and Fuhrman 2000).

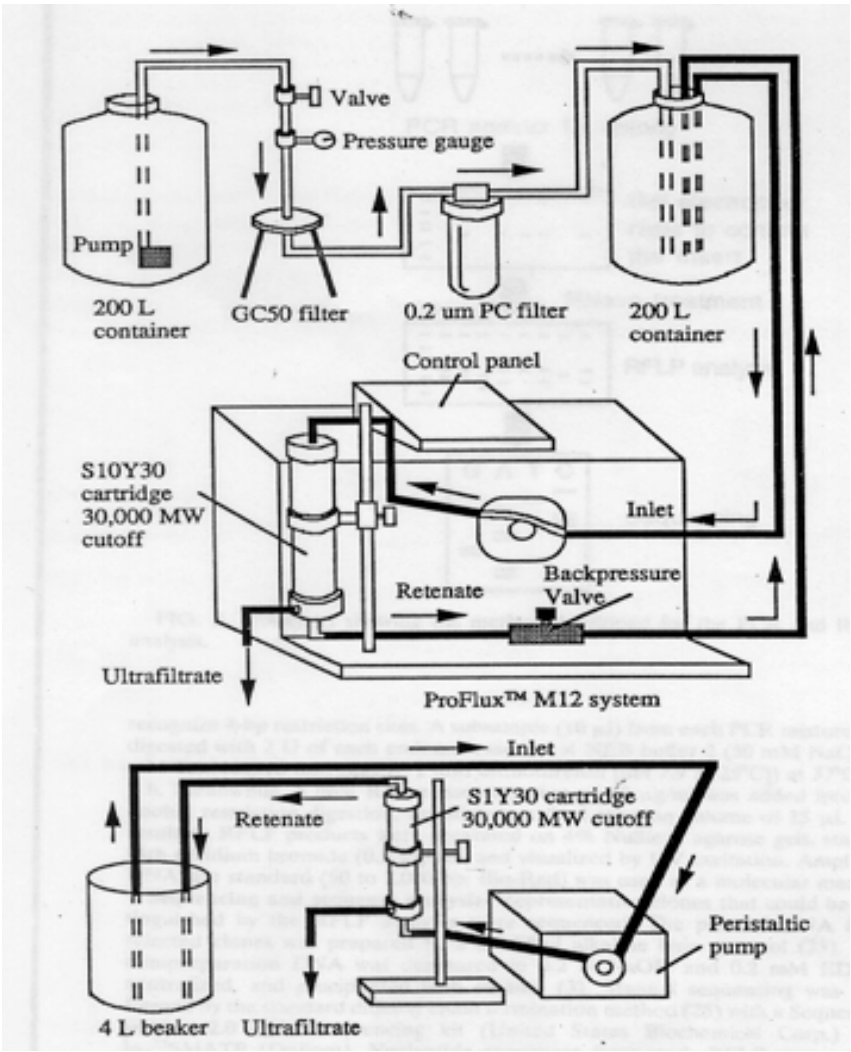

Figure 2.1 - (Suttle 1999) This figure illustrates various ultra-filtration systems used for concentrating natural virus communities from seawater. Top: Approximately 200 liters of seawater is pre-filtered through GC50 glass fiber and 0.2 μm-pore-size polycarbonate (PC) filters. The viruses in the filtrate are then concentrated into 1 to 2 liters by ultra-filtration with a S10 cartridge mounted on a Amicon ProFlux M12 system. Bottom: In some cases, the viruses were concentrated into a smaller volume (100-200 ml) by a second step ultra-filtration using a S1 cartridge and a peristaltic pump.

Culture Dependent vs. Culture Independent Approaches to Virus Studies

In order to research the presence of viruses in Yellowstone's high temperature environments a methodology had to be chosen. Most of what's known about virology today comes from studying virus-host interactions with viruses that are replicating in culture in a laboratory environment. Far less research has addressed viruses collected directly from the environment as described above in marine environments. The two approaches are fundamentally different in nature and are suited to answering different scientific questions. The first approach offers the fundamental advantage of exactly what viral host is being studied, and is better suited to answering biochemical questions such as how the virus directs the host's replicative machinery. Another distinct advantage of this method is that experimental parameters can be readily replicated. That is, the host may continue to produce the virus over time if the infection is lytic in nature another host culture can be infected with the virus and the process repeated. However, one drawback of this host-dependent approach is not knowing if having the host organism in culture alters the behavior either of the host or the virus, and if it does to what degree?

The host-independent approach of collecting viruses directly from the environment is more suited to answering ecological questions such as what viruses are present in any given environment and how that might change over time. However, this method will reveal little of virus-host interactions, and does not lend itself to knowing which virus belongs to which host. Another potential draw back is that if the biomass of the environment is very low (unlike the trophic zone in the ocean) it might increase several fold the difficulty of collecting detectable amounts of virus. Additionally there

may be other particulate present that masks the virus. This may be because the other particulate is of similar size to the virus thereby masking it, or there may be larger charged particles such as clays that the virus sticks to making filtering impossible. Another difficulty could be that virus particles may be amorphous in nature and not easily distinguishable under the TEM from other particulate present such as cellular debris.

Host Independent Survey (Filter Concentration) Method

A good deal of effort was put forth in the preliminary stages of this study to detect and isolate viruses directly from the environment by filter concentration. This effort met with little success and a summation of the process is presented in Appendix A.

Host Dependant Survey (Culturing Sulfolobus) Method

In reviewing the relatively small body of literature that encompassed discovery and characterization of viruses, plasmids, and other genetic elements of thermophilic and hyperthermophilic archaea, it was evident that a host-dependant approach in this pursuit was feasible. The procedures and results emanating from Zillig's lab at the Max Planck Institute in Martinsried, Germany, outlined a successful strategy for establishing host cultures in the lab that produced viruses and VLPs from thermal environments in Iceland and Italy (Zillig et al. 1998; Zillig et al. 1996). Their methodology consisted of sampling spring water and mud holes in solfataric fields in two manners. Briefly outlined, they would dip into the features with a large plastic beaker at the end of a long handle. After recording the pH and temperature of the sample, the pH was adjusted by the careful addition of $CaCO_3$ to a value of 5.4. After a majority of the mud settled out of the

adjusted samples they were processed in two ways. In the first method an environmental sample was injected with a syringe into 25 ml vials containing 160 kPa of CO_2, 1 kPa of H_2S, 0.2 g of elemental sulfur, and 20 μg resazurin as an oxygen indicator. If samples were slightly aerobic this was adjusted by the drop wise addition of water saturated with H_2S, then stored at ambient temperature until returned to the lab, at which point they were stored at 4°C. In the second method, 80 ml of the supernatant was filtered through small replaceable 0.22 μl Millipore filters fitted on the end of 10 ml syringes. The filtrated were then injected into 100 ml serum bottles containing 10% PEG 6000, and 1M NaCL for virus precipitation. The filters themselves were treated by blotting excess moisture from the empty side of the filters and placing them in 8 ml of salt base Brock's medium containing 20%w/v DMSO. Anaerobicity of these samples was assured by placing them in flasks with rubber stoppers and aluminum caps, and injecting H_2S-saturated water.

Back in the lab all potential host organisms were cultured by adding one g/l tryptone to the salt base Brock's medium with elemental sulfur and gassing with 5% CO_2. These cultures were grown at 80°C in erlenmeyer flasks equipped with extended 20 cm necks to avoid evaporation.

These approaches yielded significant results for this lab in the following ways. In screening the samples that were treated with PEG and NaCl, a variety of forms resembling known viruses were seen. Some of these had a roughly icosahedral head accompanied by a tail, some were rod shaped (like TMV), and others had a characteristic lemon shape with a variable size range. Many of these particles were attached to rod-

shaped cells that looked like *Thermoproteus*, and some were associated with *Sulfolobus*-like cells. Most particles, though, could not be associated with any particular host using this method.

After growing successful enrichment cultures they were able, in one case to, isolate 44 heterotrophic, and 53 autotrophic pure cultures by using extinction dilution and plating methods. All of the heterotrophic cultures were identified as either resembling S. *solfataricus* or S. *islandicus* through DNA restriction pattern analysis, and three different types of autotrophs were distinguished in the same manner. The closest match to a described species for the autotrophs was *Desulfurolobus ambivalens.*

Zillig and colleagues were able to detect several plasmids by separating covalently closed circular DNA (cccDNA) from chromosomal DNA on CsCl density gradients in the presence of ethidium bromide (Zillig et al., 1996). These plasmids ranged in size from 5.5-7.5 kb and copy numbers in cells ranged from 15 to greater than 50.

Screening of the cultured isolates with the TEM identified the presence of several virus types in this study. These included the 40 X 400 nm stiff rod SIRV, the filamentous 40 X 800 nm SIFV, and the ubiquitous 60 X 90 nm lemon shaped SSV viruses that they observed in a majority of their cultures. They also saw a filamentous virus in one of their autotrophic culture named “type D” which is morphologically similar to SIFV, which was dubbed DAFV for *Desulfurolobus ambivilens* filamentous virus.

Yellowstone contains numerous solfataric fields similar in temperature and pH range to those investigated in Iceland, Japan, and Italy that had were home to numerous

thermophilic viruses as outlined above. We reasoned that the approach pioneered by the Zillig lab, if adopted in YNP, would produce a survey of viruses similar in diversity and breadth to that which they had discovered.

With this in mind we endeavored to produce a survey of viruses in acidic thermal environments found in the YNP. The basic approach was modeled upon the methods outlined above that had been so successful in thermal fields abroad. This was to sample solfataric fields in a similar manner, bringing those samples back to the lab, and culturing probable hosts of the viruses we were hoping to survey. Once isolated cultures were established we planed to screen them for extra chromosomal elements and viruses as had been done previously.

Materials and Methods

Endeavoring to learn the growth and isolation techniques that had been fruitful in the Zillig lab we traveled to the Max Planck Institute in Martinsried, Germany and were kindly accommodated by Wolfram Zillig, as well as David Prangishvilli and Ken Stedman. Here we learned the basics of collecting *Sulfolobus* from acidic thermal features, growing it in the lab, and isolating virus from the ensuing cultures as described above. With this working knowledge we returned home and began a successful survey of thermal viruses in the Yellowstone National Park.

Sampling Locations

Sampling sites were selected that had an appropriate range of temperature (70-90°C) and pH (2.0-4.0) for *Sulfolobus* growth. Eight different thermal sites within YNP

were selected for sample collection over a 12-month period (Figure 2.2, Table 2.1). Hot springs, mud pots, and soils were included in this sampling distribution.

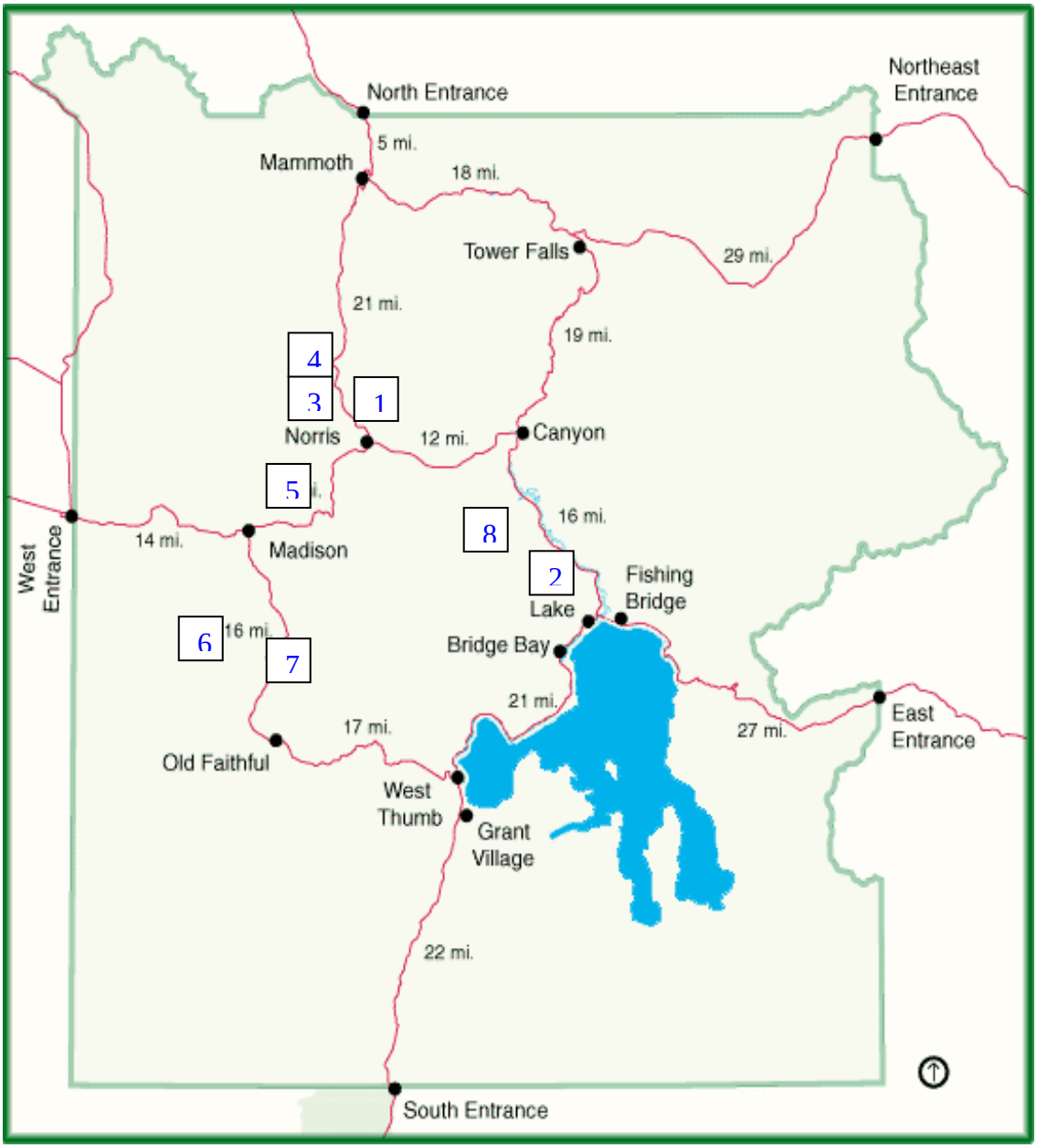

Figure 2.2 - Map of YNP with all of the sampling locations marked in blue numbers. Numbers correspond to locations in Table 2.1.

Table 2.1 - YNP thermal features selected for sample collections that correspond with the positions marked on the YNP map (above). Positions are UTM type projections, Datum type- NAD 83, Zone 12

	X Position	Y Position
1. Amphitheater Springs	521734.7	4960405.6
2. Mud Volcano	544849.3	4941373.9
3. Norris Basin	522734.3	4953262.0
4. Nymph Lake	521477.2	4955457.2
5. Gibbon Basin	518444.3	4949627.8
6. Fountain Flats	513307.2	4934500.0
7. Rabbit Creek	514867.9	4929918.2
8. Crater Hills	540943.3	4944686.8

Establishing Enrichment Cultures

Two different methods were used to establish *Sulfolobus* enrichment cultures. In the first method we proceeded much as the Zillig lab, by dipping into the features with a large sterile plastic beaker attached to the end of a long lightweight sampling pole. After recording the samples pH and temperature the pH was adjusted by addition of $CaCO_3$ to a value of 5.4. This must be done with care as the reaction proceeds slowly and the pH will creep up for some time after the addition of the $CaCO_3$. After a majority of the mud and clays settled out of the adjusted samples they were processed in two different ways. In the first method they were injected with a 10 ml syringe into 60 ml vials containing 160

kPa of CO_2, 1 kPa of H_2S, 0.2 g of elemental sulfur, and 20 μg resazurin as an oxygen indicator. If samples were slightly aerobic this was adjusted by the drop wise addition of water saturated with H_2S, then stored at ambient temperature until return to the lab at which point they were stored at 4°C. In the second method, water, mud, and soil samples from the thermal sites were collected directly into sterile 50 ml centrifuge tubes (with a long handled pole when necessary). All samples were then transported back to the lab at ambient temperature and stored at 4°C.

Environmental samples were inoculated into a salt-base media similar to that described by Zillig (originally developed by Brierley and Brierley) used for heterotrophic, aerobic and anaerobic autotrophic, and mixotrophic enrichment (Brierley and Brierley 1973; Zillig et al. 1994). Media contained per l 3g $(NH_4)_2SO_4$; 0.5 g KH_2PO_4 X 3 H_2O; 0.5 g $MgSO_4$ X 7 H_2O; 0.1 g KCl; 0.01 g $Ca(NO_3)_2$ X 4 H_2O and supplemented with the rare metals – 1.3 mg $MnCl_2$ X 4 H_2O; 4.5 mg Na_2Br_4O7 X 10 H_2O; 0.22 mg $ZnSO_4$ X 7H_2O; 0.005 mg $CuCl_2$ X 2 H_2O; 0.03 mg $NaMoO_4$ X 2 H_2O; 0.03 mg $VOSO_4$ X 5 H_2O, 0.01 mg $COSO_4$ X 7 H_2O; Wolin's vitamin mixture, 0.7glycine for buffering, and adjusted with half concentrated sulfuric acid to pH 3.1.

Enrichment cultures were established from the liquid environmental samples by withdrawing 1ml with a sterile syringe from the sealed vials and inoculating into erlenmeyer flasks equipped with 20 cm necks to avoid evaporation. Cultures from soil and mud samples were established by adding one gram of the environmental sample directly to the long-necked flasks containing the same salt-base media. The flasks were then incubated at 80°C in a shaking water bath (supplemented with mineral oil or

Polyetheleneglychol (PEG) to abate evaporation) for up to 12 days and monitored for growth.

Figure 2.3 - Top left: A pH/temperature probe is immersed in a small hot pool in order to determine if it fits the parameters required for probable thermal viral hosts. Top right: Aerobic sampling method of dipping directly into the thermal feature with a sterile 50ml tube to collect the sample. Bottom left: Sterile syringes and prepared serum vials used for direct injection in the anaerobic sampling method. Bottom middle: Modified Erlenmeyer flasks used to prevent evaporation of growth media for the collected samples at the required elevated temperatures. Bottom right: "Artificial hot springs" employed at the lab to keep the enrichment cultures at the desired 80° C. This is done by filling temperature controlled liquid shakers with either mineral oil or a PEG solution to avoid rapid evaporation of the insulating viscous liquid required to keep the cultures at a constant high temperature.

16S rDNA Analysis (performed by Jamie Snyder Doctoral Student in the Young Lab)

Total DNA was extracted from representative enrichment cultures from each of the eight sampling sites using previously described protocols (Stedman et al. 1999a). The 16S rDNA gene was amplified from the genomic DNA by standard PCR protocols using archaeal specific primers 0023a forward (CTCCGGTTGATCCTGCC) and 1492 reverse (GGTTACCTTGTTACGACTT) (Achenbach and Woese 1995). The resulting PCR products (approximately 1500 bp in length) were subsequently cloned into pCR2.1 using Topo-TA cloning protocols (Invitrogen). The complete 16S rDNA inserts were sequenced by Big Dye Termination protocols using an ABI 310 or 3700 automated capillary sequencer. Multiple independent clones from each site were sequenced and the complete sequence for each clone was determined. The resulting sequences were compared to sequences in the Ribosomal Data Project (RDP) at Michigan State University.

Virus Detection Methods

Two methods were employed for detecting the presence of virus in cultured samples. In the first method enrichment cultures were subjected to a plaque and or growth inhibition assay by spotting 1-2 μl of the culture directly onto lawns of *S. solfataricus* strain P1; *S. islandicus* REN1H1; HVE10/4; *S. acidocaldarius* DSM639 and two *Sulfolobus* isolates from YNP. *S. islandicus* culture REY15/4 (containing SSV2; (Arnold et al. 1999a) was used as a positive control. Plates were incubated for 2-3 days at 80°C and scored for plaque formation or growth inhibition zone.

In the second method all enrichment cultures were processed for direct visualization with the TEM. Enrichment cultures were grown to late stationary phase (10 days at 80°C) and the stressed by leaving them for two days at 22°C (bench top). Host cells and larger particulate were removed by centrifugation at 6,000 x/g for 10 minutes. The supernatant was filtered through Acrodisc PF 0.8/0.2 μm filters (Gelman Laboratory) and remaining particles pelleted by centrifugation at 100,000 x/g for two hours. Pellets were re-suspended in 20 μl of ddH_20, stained with 1% uranyl acetate, and examined with a Zeiss 100CA transmission electron microscope.

Transforming *Sulfolobus* with Viral Nucleic Acid

Nucleic acid from purified virus preparations was transformed into *Sulfolobus* strain P1 as described previously (Schleper et al. 1992). Fifty milliliters of *S. solfataricus* strain P1 was collected after it had been grown to an optical density (OD) of .01-.03 (usually 2-4 days) and then put on ice for 15 minutes. The *S. solfataricus* strain P1 was then centrifuged at 3 K for 10 minutes and the supernatant gently poured off the cell pellet. The cell pellet was then re-suspended in a 20 mM sucrose solution and spun again for 10 minutes at 3 K. The process was repeated by re-suspending the resultant pellet in 25 ml of the sucrose solution and then spun again in the same manner. This was again repeated with a re-suspension in 12.5 ml 20 mM sucrose and again in 1 ml. Finally, the cell pellet from the 1 ml solution was re-suspended in two times the starting OD of the P1 (i.e. if the original OD was .16 then re-suspend in .32 ml). This cell suspension was iced for 15 minutes to an hour.

Viral DNA was then added in conjunction with 50 μl of the cell suspension to an electroporation cuvette. The cuvette was then placed in a BTX expediential decay electroporater and electroporated at 1.5 KV and 129 ohms. The electroporated cells are then removed from the cuvette and placed in 1 ml of YS media (Zillig media supplemented with 0.1% yeast and 0.1% sucrose), warmed to 80°C, and then incubated for one hour. This was then spotted (1-3 μl) on a growing lawn of *Sulfolobus* strain P1. The remainder of the electroporated cells were diluted into 25 ml of YS media and incubated for two days at 80°C while shaking. This culture can be prepared as outlined above to look for virus with the TEM.

Viral Genomic Library Construction

Total DNA libraries were constructed from virus preparations (as described above) of three separate enrichment cultures, each containing a single virus particle morphology. Total DNA was extracted from each sample by proteinase K, SDS, and phenol method (Sambrook et al. 1989); digested with either *Eco*RI, *Sac*II, or *Bam*HI restriction endonucleases; ligated into pBlueScript II SK+; and transformed into XL2 Blue MFR' *E. coli* (Stratagene). Recombinant plasmids were screened for insert size. Inserts of different size from each library were sequenced with M13 forward and reverse primers using Big Dye Termination (Applied Biosystems) on ABI 310 or 3700 automated DNA sequencers. Sequences were compared to known *Sulfolobus* viral sequences and multiple sequence alignments were performed using GCG and BLAST (Altschul et al. 1997).

Results

Enrichment Cultures

Yellowstone National Park is estimated to have more than 10,000 geothermal features. Many of these thermal features are acidic and are likely to be favorable environments for *Sulfolobus* species and their viruses. A total of 183 hot springs, mud pots, and soils at eight different sites within YNP were sampled in this study. A total of 51 enrichment cultures, representing all of the geographically distributed sites, were successfully established (Table 2.2).

Table 2.2 - Locations of sampling sites, as well as the temperature and pH of the features. Numbers of samples taken at each site is recorded along with the number of cultures established from theses samples. Positions are UTM type projections, Datum type- NAD 83, Zone 12

Site Name	X Position	Y Position	Temp.	pH	# Of Samples Collected	# Of Cultures Established
1. Amphitheater Springs	521734.7	4960405.6	74-85	1.0-3.5	34	5
2. Mud Volcano	544849.3	4941373.9	70-85	2.0-4.0	25	8
3. Norris Basin	522734.3	4953262.0	70-84	1.5-4.5	60	5
4. Nymph Lake	521477.2	4955457.2	74-84	2.0-4.5	16	7
5. Gibbon Basin	518444.3	4949627.8	70-83	1.8-3.5	9	8
6. Fountain Flats	513307.2	4934500.0	74-87	3.0-4.1	8	4
7. Rabbit Creek	514867.9	4929918.2	72-92	2.9-3.9	16	11
8. Crater Hills	540943.3	4944686.8	73-81	2.5-3.6	15	3

Enrichment cultures typically took from three to ten day to show signs of growth if they were successful. They were monitored daily for signs of turbidity and a characteristic “dirty sock” smell indicating an active *Sulfolobus* culture. The appearance of the turbidity varied from the clear yellowish appearance of the media by itself to a turbid creamy appearance that ranged in color from white to a more yellow hue. This coloration was effected by the color and quantity of sample added to the environmental soil and mud samples, as one would expect. Cultures often ranged from values of 0.3-0.8 abs units at 600 nm when checked in a light spectrometer at the peak of their growth phase. No matter what the appearance though, the characteristic smell of *Sulfolobus* was always present in a successful culture.

When viewing a growing culture under the light microscope at 1,000 X magnification the *Sulfolobus* cells appeared dark and cocoid in nature and moved fairly rapidly across the field of view in a tumbling fashion. Depending on the growth phase and density of a culture you can usually see from just a couple of cells to hundreds in one field of view (in the range of 10^5–10^8 cells per ml). The vivacity of the tumbling motion is a good indicator of the health of a culture. If a culture was left out bench top for several days you could still see numerous cells, but they would not display their rapid motion. Cultures could normally be passaged by simply withdrawing 1.0 ml from an active culture and inoculating fresh media in a new long neck erlenmeyer flask, or by pouring off the bulk of a culture for processing and replacing that volume with fresh media. If a culture was not to far into stationary phase (after 7 days) it would usually passage well in this fashion and be back into a mid-log growth phase within two days.

16S rDNA Analysis

The sequences of the 16S rDNA genes were determined from selected enrichment cultures representing seven of the eight geographically distributed sites. A single 16S rDNA sequence was detected from each of the enrichment cultures tested, suggesting that a single archaeal species dominated each enrichment culture. Alignment of the 16S rDNA sequences from these enrichment cultures demonstrated that six of the seven sequences were identical. Comparison of these 16S rDNA sequences with those in the RDP showed that these sequences were almost identical (>99%) to the 16S rDNA gene sequence of *S. solfataricus* previously determined for isolates from Italy. However, one enrichment culture established from Gibbon Basin had a 16S rDNA gene sequence that was most similar to *Metallosphera prunae* at 93% (clone library construction and sequence analysis performed by Jamie Snyder).

Detection of Viruses and Virus-like Particles

Six distinct virus and virus-like particle morphologies were detected in 22 of the 51 enrichment cultures (Table 2.3). The virus or virus-like particle morphologies could be separated into two categories; those particles that have *morphologies that appear similar* to viruses isolated from *Sulfolobus* species from Japan and Iceland (SSV, SIRV, and SIFV-like particles shown below); and those particles that have *morphologies that have not been previously identified* with viruses from *Sulfolobus* (32 nm spheres, double-tailors, and STIV particles shown below). Twelve of the enrichment cultures contained

multiple virus particle morphologies (figure 2.4). These enrichment cultures with mixed viral morphologies are not included in Table 2.3.

Table 2.3 - This table shows the characteristics YNP viruses isolated along with the number of enrichment cultures each type was found in. The upper three entries resemble viruses already described by the Zillig Lab in other thermal areas around the world, and the lower three are virus-like morphologies not previously described.

Class	Morphology	Dimensions	# Of (+) cultures
SSV-Like	Lemon Shaped	Var. 40 x 80 nm	18
SIRV-Like	Rods	20 x 500 nm	6
SIFV-Like	Filaments	30 x 1000 nm	2
Unique	Spherical	30 nm	7
Unique	Head & Tail	100 x 800 nm	2
Unique	Spherical	52-70 nm	2

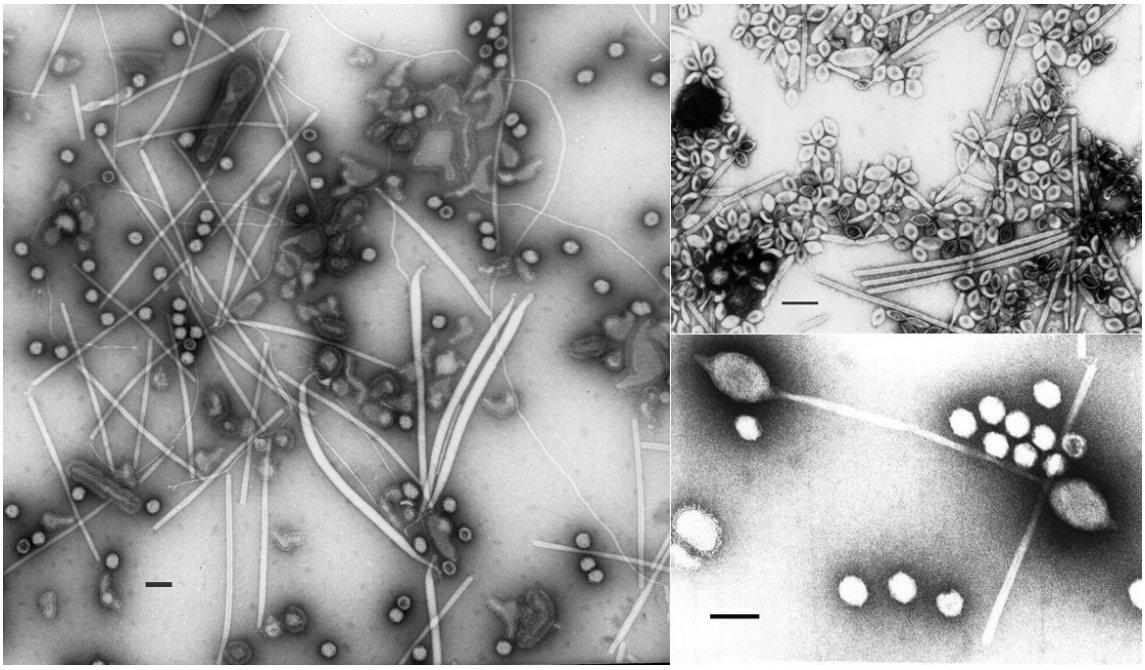

Figure 2.4 - Above left: A TEM from a culture inoculated with material taken from the Mud Volcano area that contains particle morphologies that are similar to SIFV, SIRV, as well as a previously unseen 70nm icosahedral particle. Upper right: Image is taken from the Rabbit Creek area and contains particles that look like SSV and SIRV, as well as some particles that may be related to SSV but are much larger in both dimensions or stretched into a longer and thinner configuration. Lower right: Image is also from the Rabbit Creek area and contains the 70 nm icosahedral particles and spindle-shaped particles with elongated tails on one end that had not been described before this study.

SSV - Morphologies that Appear Similar

The most common virus morphology represented in 18 out of 46 of the enrichment cultures (Table 2.3), appeared to be members of the SSV family (*Fuselloviridae*). The virus particles were similar in size and shape to SSV-1 isolates from Japan (Stedman et al. 1999b) and SSV-2 isolates from Iceland (Schleper et al. 1992).

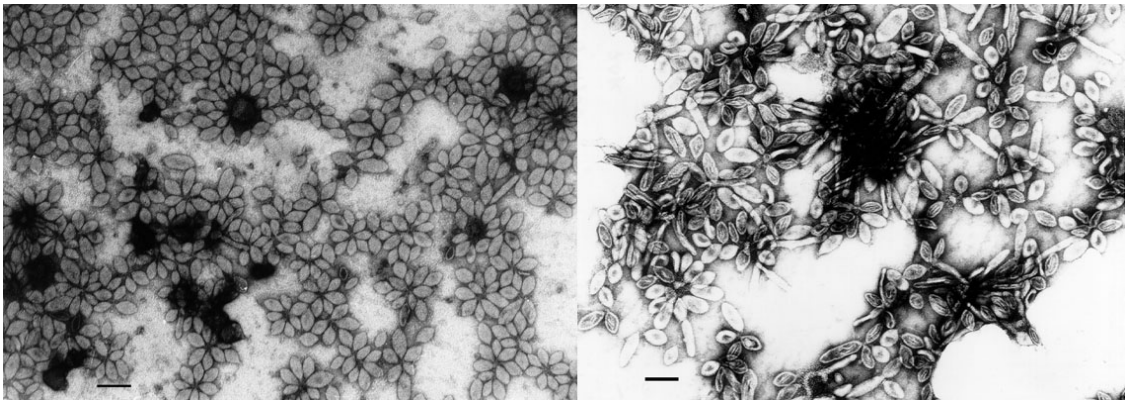

Figure 2.5 - The 90 x 60 nm spindled-shape particles are often observed in rosette structures as is seen in the image above left. The image above right exemplifies the plasticity of the SSV morphology showing long, tall, skinny, and fat morphologies.

These particles contain a ccc dsDNA genome of approximately 15 kb. BLASTTN analysis of a limited DNA sequence determined for a YNP isolate (SSVY) shows distinct similarity to SSV1 and SSV2. No other homologues were detected in the public databases. Sequence alignment of the coat protein genes and their predicted products from SSV1, SSV2, and the YNP isolate revealed only limited identity (Figure 2.6).

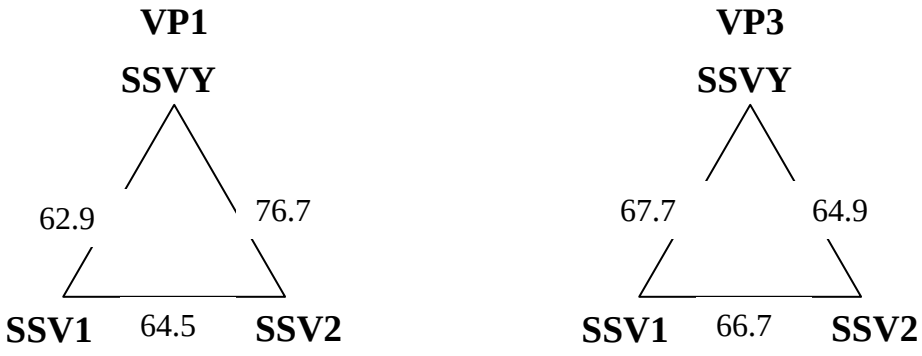

Figure 2.6 - Amino Acid Sequence Identity Comparison of VP1 and VP3 Coat Proteins in SSVY, SSV1 and SSV2.

Between the three isolates there was approximately 65% identity at the nucleotide level and 71% at the amino acid level. These genes appear to be the most highly conserved in our analysis. While this limited analysis demonstrates that the YNP isolates are related to the Japanese and Icelandic isolates, it appears to be a distant relationship. The YNP isolates appear to be as distinct from the Icelandic and Japanese isolates as the latter are from each other.

SIRV – Morphologies that Appear Similar

The second virus particle morphology is characterized by rigid helical 25 nm X 900 nm rods with characteristics similar to SIRV1 and SIRV2—two members of the *Rudiviridae* virus family (Prangishvili et al. 1999a). The YNP SIRV-like particles were observed in six of the 46 enrichment cultures and were often found in association with YNP SSV-like particles.

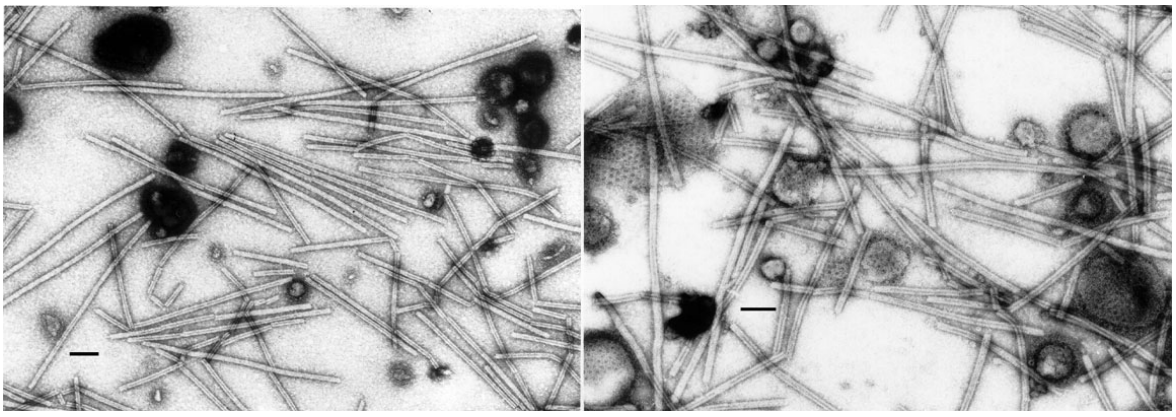

Figure 2.7 - SIRV particles were isolated from the Rabbit Creek and Nymph Lake areas. The larger circular crenulated forms seen on the right are remnants of S-layer from the host cells.

The YNP SIRV-like particles possess dsDNA genomes 33-36 kb in length as estimated from viral DNA library constructs. Sequence analysis indicates limited sequence similarity with SIRV1 (Prangishvili, *personal communication*). The highest degree of similarity found in the analysis of 8 kb of sequence was 78% identical over a 393 nt. (clone library construction and sequence analysis performed by Jamie Snyder)

SIFV -Morphologies that Appear Similar

A third virus particle morphology is characterized by a long flexible rod with characteristics similar to those of the *Lipothrixvirus*, SIFV, previously isolated from Iceland (Arnold et al. 2000b). The YNP SIFV-like particles were observed in two of the 46 enrichment cultures and are 50 nm X 900-1500 nm flexible rods with apparent attachment fibers at both ends. Partial sequence analysis (8.5 kb of non-contiguous sequence) showed a 45-65% range of identity with SIFV.

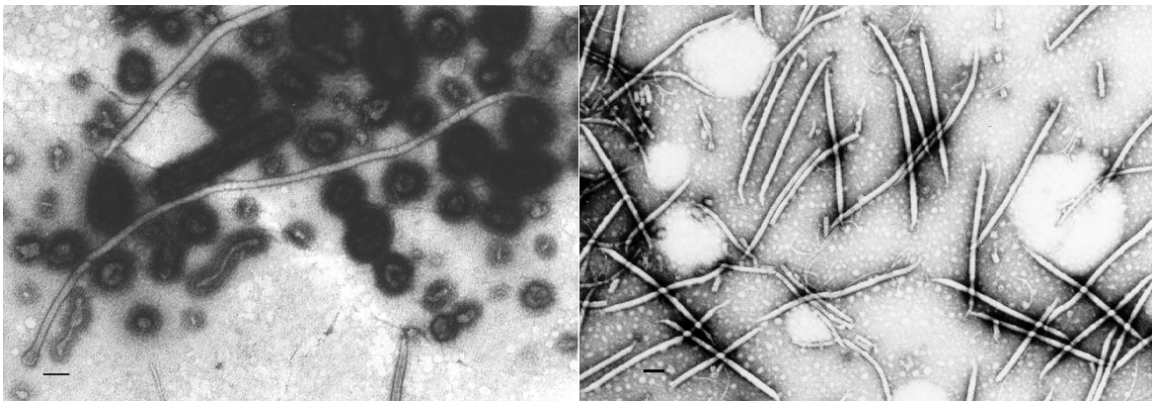

Figure 2.8 - TEMs of SIFV-like particles isolated in YNP.

32nm Spheres - Morphologies Not Previously Identified

Thirty-two nanometer icosahedral virus-like particles were observed in seven of the 46 enrichment cultures from two separate sample sites (Table 2.3). Interestingly this and only one other virus-like particle were isolated from soil samples. These virus-like particles do not appear to be enveloped. They do not form plaques on any of the tested *Sulfolobus* lawns, but appear to be induced by stressing the enrichment cultures.

A couple of points of interest in relation to the 32 nm spherical particles are that they were found in cultures that came from the same types of samples as the 70 nm icosahedral particles discussed next. Both particle morphologies were found in cultures from Rabbit Creek samples, and steaming soil sites in the Mud Volcano area. These are the only two viral morphologies that were seen in cultures from the steaming soil samples, but there were several other viral morphotypes isolated from the Rabbit Creek area besides these particles. Also the 32 nm particles were often seen in cultures from a single colony isolate that produces the 70 nm icosahedral particles (STIV).

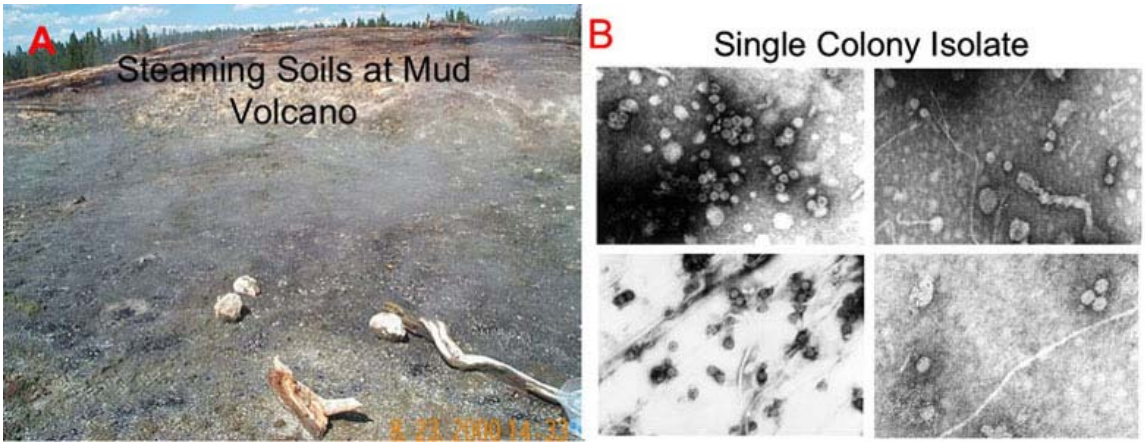

Figure 2.9 - (A): One of the steaming soil sites in the Mud Volcano area from which the 32nm spheres were isolated. This morphology was also seen in cultures that were

inoculated from samples taken from the Rabbit Creek area. (B): TEM's taken from a single colony isolate of one of the cultures producing this virus-like morphology

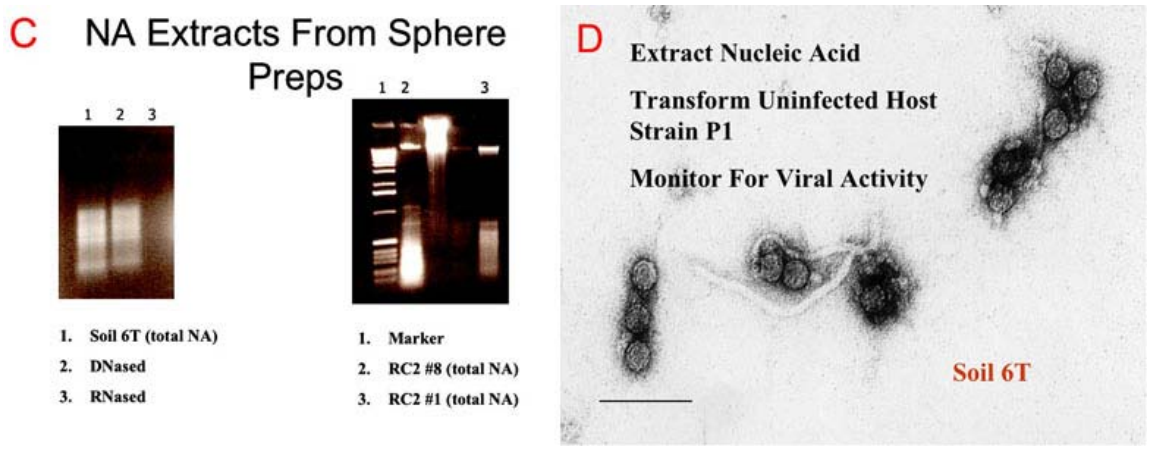

Figure 2.9 continued - (C): an agarose gel of a nucleic acid extract from a prep of the 32nm spherical particles. Lane 1 is the untreated material, and lane 2 and 3 are treated with DNase and RNase respectively. (D): A representation of the process to show that 32nm particles are infectious. Nucleic acid extracted from purified virus was transformed into uninfected Sulfolobus P1 and then the culture was monitored by TEM for the presence of virus.

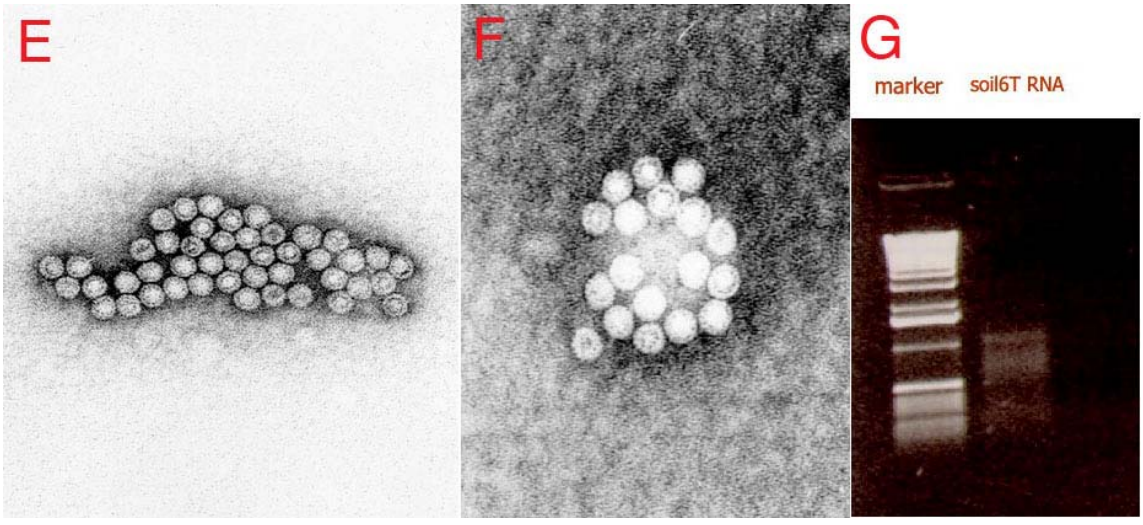

Figure 2.9 continued - (E & F): TEM showing the presence of 32nm spherical particles present in a P1 culture that was transformed with the nucleic acid of purified virus. (G): Marker in left lane with nucleic acid extracted from a viral prep of the transformed culture treated with DNase next to the marker, and the same prep treated with RNase immediately to the right of that.

Double-Tailers – Morphologies Not Previously Identified

A unique morphology was observed in two of the 46 enrichment cultures (Figures 2.10 and 2.11). These virus-like particles have either a large spindle-shaped center with appendages at each end of the central body, or a spindle-shaped end with the appendage forming a tail-like structure. The dimensions of the central body are 100 nm x 200 nm. The appendices have approximate dimensions of 50 nm x 100-200 nm. The overall tip-to-tip length is about 400 nm. These virus-like particles are often observed in rosette-like structures and may be associated with the *Sulfolobus* S-layer. The morphology of these virus-like particles does not resemble any known virus.

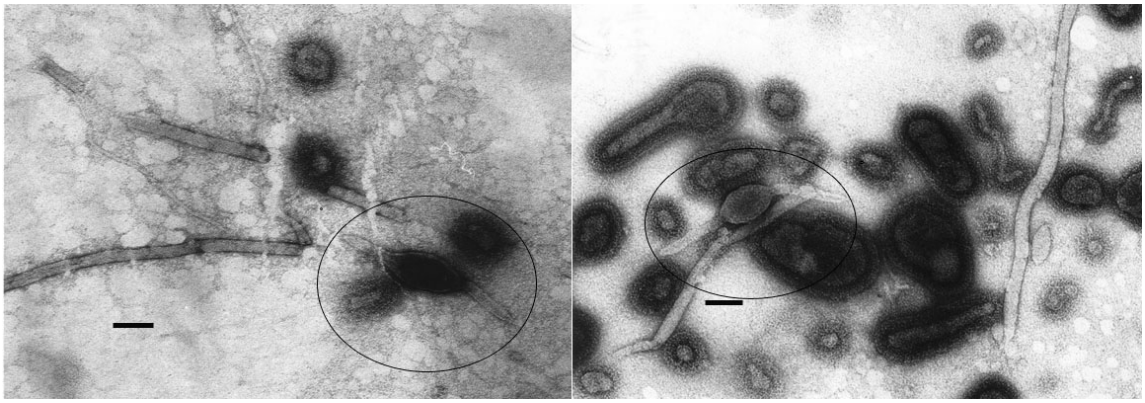

Figure 2.10 - The two images above show particles isolated from the Ragged Hills area in Norris. The double-tail morphologies are intermixed with SIFV-like particles suggesting that they may be a variation of that morphology with exaggerated bulges either in the middle or at one end of the particle.

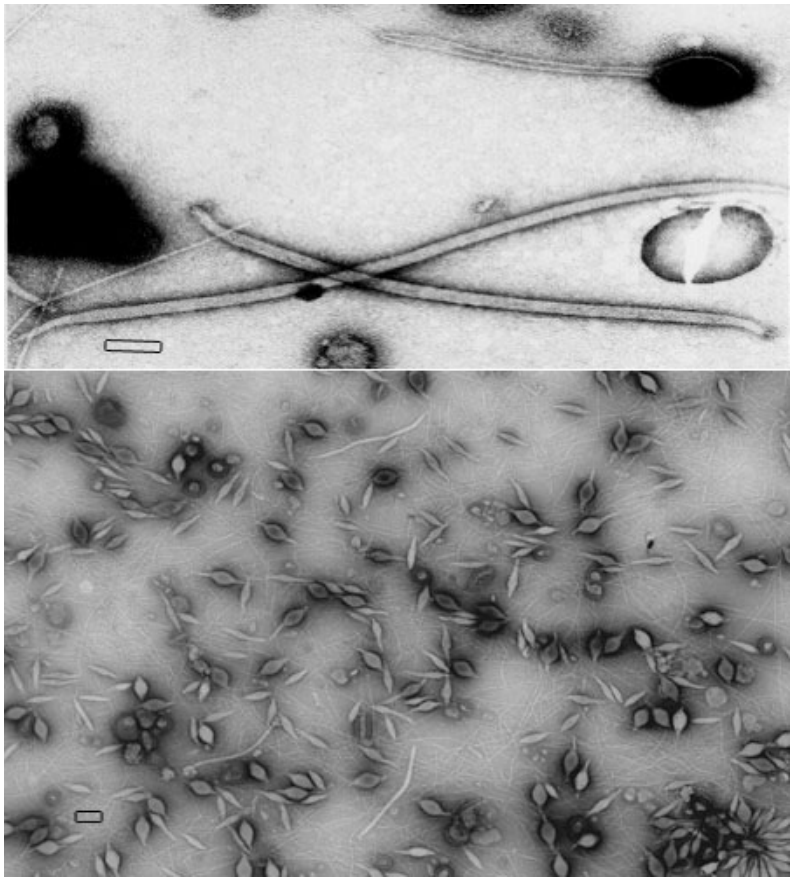

Figure 2.11 - The top portion of the above panel shows an alternate head and tail morphology, and the bottom image taken from the Rabbit Creek area shows a substantially higher concentration of the particles than was seen elsewhere. You can also see the particles tendency to form rosettes as SSV does.

STIV - Morphologies Not Previously Identified

A third type of particle was detected in three of 46 enrichment cultures (Figure 2.12). The particles are 44-70 nm in diameter, non-enveloped, and appear to be icosahedral in morphology. At each of the vertices there are 10-20 nm projections with unique mushroom-shaped ends. There appear to be 12 regularly spaced projections per virion suggesting that these projections may be from the five-fold axis of the particle. This virus is discussed at length in the next chapter.

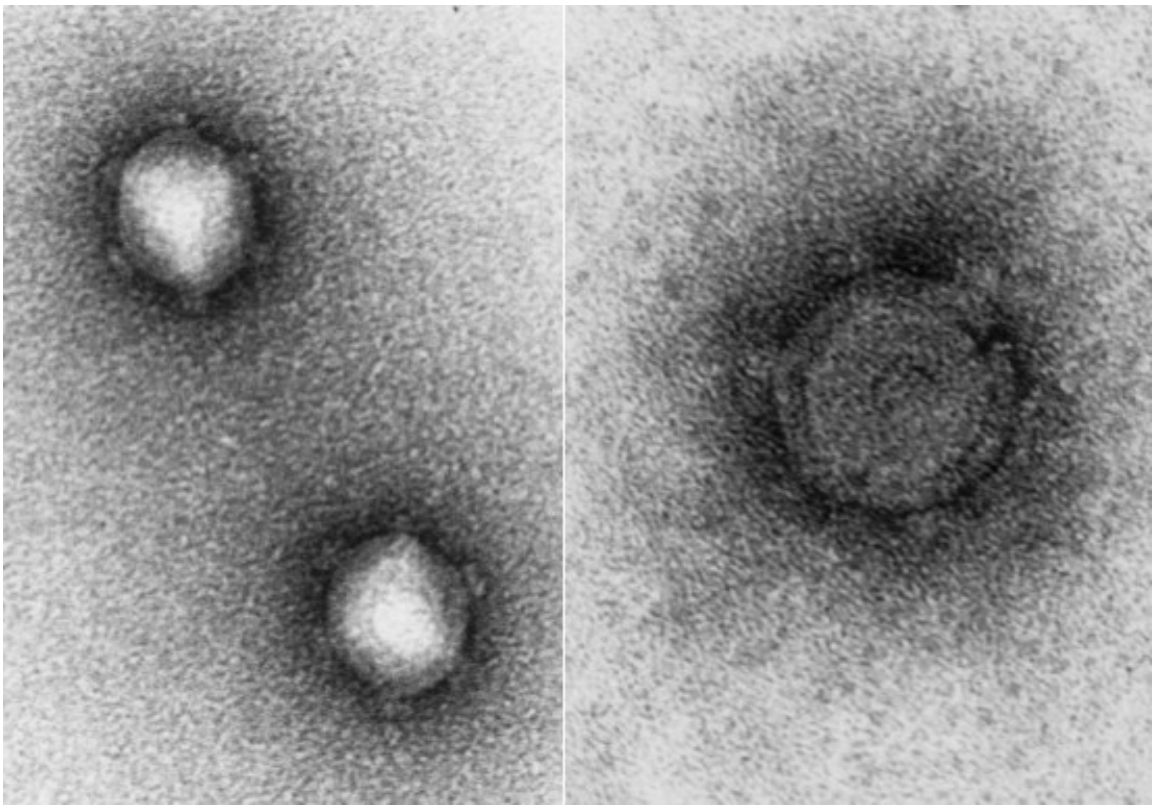

Figure 2.12 - TEMs taken by Ken Stedman of particles in a culture grown from a single colony isolate taken from the Rabbit Creek area.

Discussion

When employing a culture dependant method of virus capture such as we did here there are at least two factors that are likely to lead to the underestimation of microbial and viral diversity. The first factor is that there is always an inherent culturing bias when dealing with the hosts, which is likely exacerbated in extreme environments (Snyder et al. 2004). The "great plate anomaly," which has been recognized for some time now, states that it is likely that less than 0.1% of the organisms in any given environment can be

cultured. This likely due to many factors, one of them being the dissimilarity of culture conditions to that of the natural environment thereby inducing selective pressures.

Our observation in this study that *S. solfataricus* is the dominant organism in these environments is likely flawed due to the limit of initial sample size and primer bias of the archaeal specific primes used. Later studies in the lab showed that *Acidianus* and *Stygiolobus* were dominant in at least two of the sites we used to establish enrichment cultures (Snyder et al. 2004). Although we were able to become more proficient at establishing enrichment cultures from environmental samples as the study progressed, the limited 16S rDNA results showed that the majority of what we were culturing was *Sulfolobus*, and we were likely missing the dominant species.

Another observation that points to a probable survey bias of the viral population is that a vast majority of the enrichment cultures observed with VLPs present cured themselves of the particles within a few passages. It's unlikely that this mimics the viral population dynamics in the thermal features. Snyder et al. observed in their study using SSV to gauge culturing bias upon viral population dynamics that there was indeed a large bias imposed in this manner (Snyder et al. 2004).

This work was the first comprehensive survey of viruses and virus-like particles isolated from high temperature and acidic environments in YNP. Six different particle morphologies were found in *Sulfolobus* enrichment cultures grown at 80°C and pH 3 (Table 3.3). Virus and virus-like particles were detected in a high proportion of enrichment cultures (43%), suggesting that viruses are commonly associated with *Sulfolobus* species in YNP. Three of the particle morphologies are similar to viruses

previously isolated from *Sulfolobus* species from Iceland and Japan. Three virus particle morphologies have not been previously observed in thermal environments and some of these morphologies appear to be novel.

Despite the fact that SSV-like, SIRV-like, and SIFV-like particles in YNP have nearly identical morphologies to the *Fusellovirus* (SSV), *Rudivirus* (SIRV) and *Lipothrixvirus* (SIFV) isolated from Japan and Iceland, limited analysis of their genomic sequence indicates that they are only distantly related. This genomic diversity may reflect their long-term geographic isolation or it may be a function of adaptation to unique features of *Sulfolobus* species present in YNP. Ongoing studies are aimed at addressing the level of diversity within YNP as compared to related viruses present in other thermal regions of the world.

It is likely that this initial survey has underestimated the total diversity of viruses that exist in Yellowstone National Park's thermal environments. This study examined only a limited number of representative high temperature acidic sites and represents only a fraction of the more than 10,000 thermal features of varying pH and temperature in YNP. In addition, culturing conditions were selected based on the likelihood of supporting *Sulfolobus* growth, thereby limiting the growth of other potential virus hosts. A large number of other potential virus host organisms are known to exist in YNPs thermal sites (Barns et al. 1994).

We failed to observe virus particles by direct filtration of hot spring water using techniques that have been highly successful for detection of abundant viruses in marine environments (Suttle 1999). This may be due at least in part to the reduced total biomass

present in acidic thermal hot springs (or that the harshness of the environment selects for viruses that principally reside inside their hosts). Both of these factors would have a profound effect upon the numbers of independent virus particles available in an environment for detection by this method.

Nevertheless, the use of enrichment cultures proved to be successful for the isolation of both novel virus morphologies and those that are related to known viruses from extreme environments. The detection of novel virus morphologies suggests that there are many more thermal viruses yet to be discovered in places like YNP. The fact that the new viral morphologies identified by this work do not resemble other known viruses from either thermal or non-thermal environments suggests that their analysis will provide future insights into structure and function. It is possible that further analysis of these viruses may indicate that they represent new virus families, as did the first four families of viruses found in *Sulfolobus* (Zillig et al. 1996). This number of unique virus types isolated from a single host is truly remarkable (especially when considering the harshness of the environment).

In contrast to the diversity of virus types, there does not appear to be similar host diversity. The limited analysis of the 16S rDNA from several of the YNP enrichment cultures that were sequenced indicates that the *Sulfolobus* host species are nearly identical. The apparent contradiction of low host diversity with high virus diversity is intriguing. A thorough analysis of viruses present in *Sulfolobus* species is likely to provide unique insights into biochemical adaptations to life in extreme thermal environments. In addition, a detailed understanding of the viral replication cycle in

Sulfolobus species should facilitate the elucidation of cellular processes in Archaea and lead to a more thorough understanding of this unique domain of life.

Finally the possible connection between the 32 nm spherical particles (possibly icosahedral) and the 70 nm icosahedral virus (STIV) found in cultures from steaming soils is intriguing. The fact that the smaller 32 nm particles were detected periodically in the single isolate culture that mainly expresses the larger 70 nm particle STIV lend credence to the hypothesis that the two may be associated in some manner. It's possible that they are a result of an alternate expression of the same genome under slightly different environmental conditions, or that one is a by-product of the other. In any case further study of their probable link would likely yield insight into the fascinating phenomenon of the sheer number of alternative viral morphologies that are found in these thermal environments (Prangishvili and Garrett 2004). Many of the viral morphologies shown above appear to be somewhat plastic in nature. For instance the SSV's often appear in elongated and stretched from their most common configuration of a 60 nm X 90 nm lemon-shaped particle.

It is possible that the observation of similar viral morphotypes found in different thermal areas around the world of like temperature and pH is due to a common distant ancestry (of themselves and their hosts) in similar environmental conditions (early earth). If the wide geographic separation and present day uniqueness of these extreme environments reflects a true barrier to genetic exchange for these viruses, then the fact that they have maintained similar morphologies may reflect that these configurations convey a selective advantage. The other possibility is that they are not truly isolated from

one another and that genetic exchange does occur in some far-flung fashion such as very long distance aerosol transmission or connectivity of sub-terrain hydrology. However, studies about the diversity of the host genera *Sulfolobus* do make a case for true bio-geographic isolation (Whitaker et al. 2003). The fact that the genomes of these sister morphotypes are notably dissimilar could be due to the variable geochemical nature of the thermal features, and that the variability reflects adaptation to each unique hot spring environment.

Finally it is obvious, even to the casual observer, that crenarchaeal thermophilic virology is only in its infancy. The fact that there have only been 21 thermophilic crenarchaeal viruses described to date as compared to the approximately 5100 viruses known to infect bacteria and eukaryotes (Ortmann et al. 2006) indicates one of two things. Either there are just not many of these amazing pseudo life forms present to be studied, or we have investigated this field in a very limited manner to date. I would argue the latter. Compared to the time spent characterizing the viruses of bacteria and eukaryotes (particularly the ones involved in pathogenicity) the time spent on this endeavor has been scant.

Classic virology has shown that where there is life there are viruses, and often multiple types of viruses for each individual host species. This being the case, if there is biomass present in any given environment then you would expect to find a probable ratio of viral forms associated. As epifluorescence microscopy has recently shown that there is from one hundred thousand to a million cells per ml in YNP high temperature acidic environments (Ortmann et al. 2006), it would follow that the viral diversity and number

have yet to be scratched. It is likely that not only will we find many more unique virus types in these environments with further study, but also that elucidation of their mechanisms will yield many relevant answers to questions in biochemistry and biotechnology.

CHAPTER 3

CHARACTERIZATION OF *SULFOLOBUS TURRETED ICOSAHEDRAL VIRUS* (STIV)

Compared to Bacteria and Eukarya, very few viruses have been identified from Archaea. This chapter documents the structure of a hyperthermophilic virus isolated from an archaeal host found in YNP hot springs. The sequence of the circular dsDNA viral genome shows that it shares little similarity to other known genes in viruses or other organisms. The structure reveals tertiary and quaternary relationships in dsDNA virus capsids that span all three domains of life suggesting that these viruses shared a common ancestor preceding the division of the three domains of life more than 3 billion years ago.

Prior to this study there had been no detailed structural analyses of high temperature viruses from Archaea reported. Here we report the first structural and genetic characterization of a virus from a hyperthermo-acidophilic archaeon, and the first characterization of any sort of a thermal virus with an icosahedral structure. As mentioned in the previous chapter, spherical virus like particles measuring approximately 70 nm in diameter were seen in cultures established from samples taken both from both near the Rabbit Creek Monitor Site (Figure 3.1), and from steaming soil/mud pot samples taken in and around Little Gumper in the Mud Volcano area.

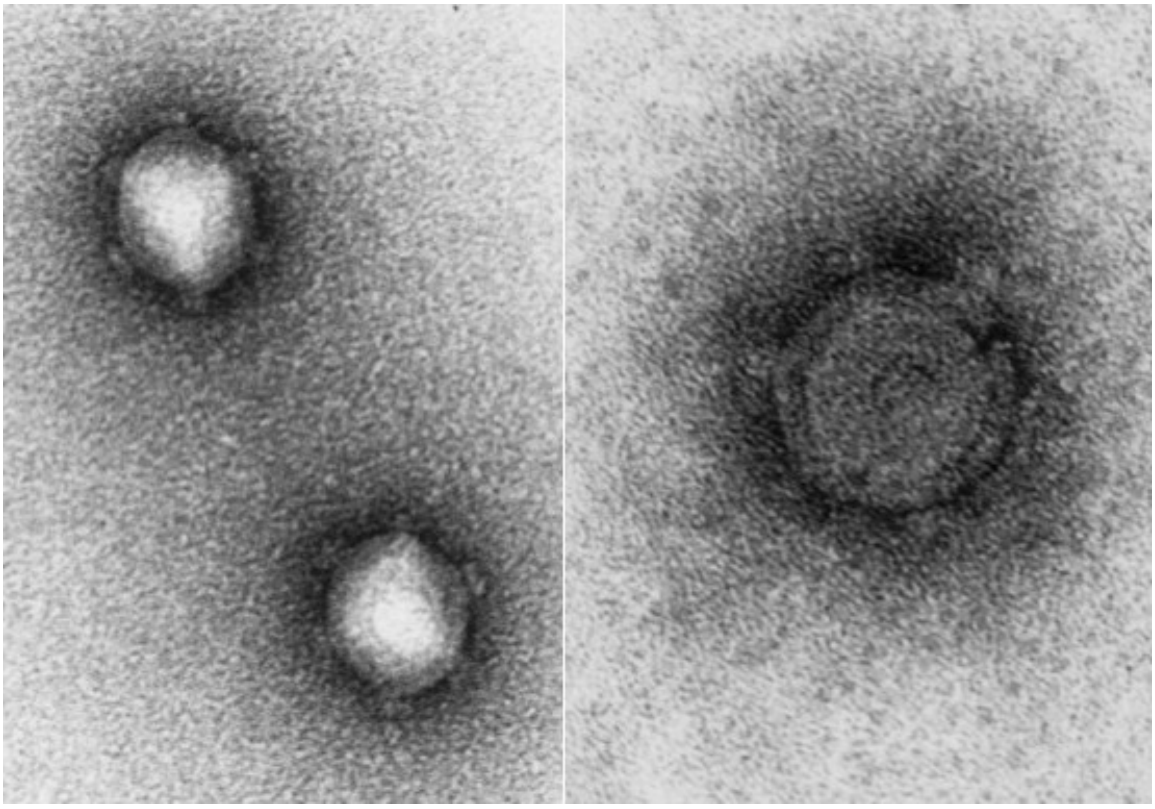

Figure 3.1 - TEMs taken by Ken Stedman of particles in a culture grown from a single colony isolate taken from the Rabbit Creek area.

Materials and Methods

In the initial characterization of this virus, a key component in the success of the effort was to produce the particle in sufficient pure quantities to perform the following: to extract homogeneous nucleic acid from the particle to begin genomic identification through restriction digests; create clone and shotgun libraries in order to sequence the genome; isolate proteins for analysis, and purify a high enough concentration of pure particle to perform cryo-electron microscopy.

The sequence of events important in the success or these efforts were; (1) the host was isolated in a single colony by Ken Stedman (Postdoctoral Fellow in the Young

lab); (2) it was determined that over time this single colony isolate did not cure itself of the virus; (3) the host cultures had to be grown in quantities allowing for sufficient production of the virus; (4) the conditions for adequate production of the virus by the host had to be determined; (5) purification of the virus from host contaminant while retaining the structural integrity of the particle; (6) demonstration that the particle would infect a known virus-free *Sulfolobus* host; (7) purification of the virus to a high enough concentration to perform cryo-EM; (8) bioinformatic analysis of the genome; and (9) structural analysis utilizing the results from the cryo-EM 2D image reconstruction.

Sampling Locations

The Rabbit Creek study site is located within the Midway Geyser Basin and contains a wide variety of thermal features such as hot pools, mud pots, and fumaroles. These features are quite diverse in appearance, having a wide range of sizes, colors, turbidity, clarity, and consistency. Pools range in clarity and color from transparent aqua marine to very murky browns, grays, and a striking tomato soup-like red. Mud pots range in color from red to white to brown, and also have a wide range of viscosities. In general, as one progresses from the outflow of the basin (approaching from the road to the west) pools range from cooler (~ 45°C), mid-sized (1-3 meters) and pH 5.5-7.0, to warmer (75-90°C) and more acidic (pH 2-4.5). Many features on the bench, which forms the high point of the basin to the east, range in size from less than a meter to several meters in diameter, and are generally very turbid browns and reds and acidic (pH 2-4.5).

The sample used to establish the enrichment culture expressing STIV was taken from a series of small acidic thermal pools (pH 2.9-3.9, 72-92˚C), slightly east of the

Rabbit Creek Monitor Site (44° 31.287′N, 110° 48.647′W), at the base of the hill as you progress upward toward the bench.

Virus Isolation

Enrichment cultures were established from the above-mentioned pool in the Rabbit Creek thermal basin within the Midway Geyser Basin. A single colony isolate of *S. solfataricus* expressing a spherical virus-like particle was established by previously described methods (Schleper et al. 1992). Briefly, 500 µl of a tenfold concentrate from a mid-logarithmic enrichment culture were mixed with 5 ml of freshly prepared, hot (70-80°C), soft overlay containing 0.3-0.4% Gelright in preheated glass tubes and poured onto pre-warmed Gelright plates containing the same growth medium. The plates were incubated at 80°C and monitored for turbid plaques (areas of growth inhibition) daily for several days. Upon the formation of plaques they were picked with sterile pipette tips, placed in liquid media, and incubated again at 80°C. These cultures were monitored for growth of the host by smell, and by direct visualization with the TEM for the presence of virus. This isolate, which produced the 70 nm spherical particle, was designated YNPRC179. The 16S RNA gene of the isolate was sequenced and compared to other thermophilic archaea to determine the genus and species of the host.

Virus Purification (Small Batch)

Isolate YNPRC179 was inoculated into medium 88 (Table 3.1), and grown as an enrichment cultures as described in the previous chapter at 80°C.

Table 3.1 - Recipe for DSM medium 88

$(NH_4)2SO_4$	1.30 G 1.31
KH_2PO_4	0.28 g
$MgSO_4$ x 7 H_2O	0.25 g
$CaCl_2$ x 2 H_2O	0.07 g
$FeCl_3$ x 6 H_2O	0.02 g
$MnCl_2$ x 4 H_2O	1.80 mg
Na_2B4O_7 x 10 H_2O	4.50 mg
$ZnSO_4$ x 7 H_2O	0.22 mg
$CuCl_2$ x 2 H_2O	0.05 mg
Na_2MoO_4 x 2 H_2O	0.03 mg
$VOSO_4$ x 2 H_2O	0.03 mg
$CoSO_4$	0.01 mg
Yeast extract (Difco)	1.00 g
Water, freshly distilled	1000.00 ml

Adjust pH to 3.0 with 10 N H_2SO_4 at room temperature.

Culture supernatants were monitored periodically for the presence of virus by direct visual observation with the TEM as described in Chapter 2. When virus titers reached maximum levels as evidenced by seeing more particles in several fields of view with the TEM than at any other time (typically 3 days), host cells were removed by centrifugation (6,000 x g for 10 minutes), and the supernatants were filtered through

Acrodisc PF 0.8/0.2 μm filters (Pall Gelman Laboratory). Filtrate was concentrated by passage through Amicon 100 membranes (Millipore Corporation) until the retained volume was approximately 8 ml. At this point the preparation was inspected for viral concentration and purity with the TEM. If the retenate was viable with at this point, it was loaded onto Cs_2SO_4 (38% wt/vol in 25 mM citrate buffer pH 3.5) and centrifuged at 38,000 x g for 20-24 hours. Virus bands were removed by carefully removing distinct bluish bands in the gradient with glass pipette. The virus bands were then placed in 14,000 mw cut off dialysis tubing, and dialyzed at 4°C against 25 mM citrate buffer (pH 3.5). Dialysis buffer (approximately 200 ml) was changed at least twice. Excess volume gained in the dialysis process was reduced by re-concentrating the virus preparation with a Microcon 100 (Millipore Corporation) when necessary. The resulting purified virus was characterized by TEM; dynamic light scattering (90 Plus Particle Size Analyzer, Brookhaven Instrument Corp.); UV/Visible spectroscopy (Agilent 8453, Agilent Technologies); and SDS PAGE.

Virus Purification: Large Batch

YNPRC179 was inoculated into 3-4 one-liter long neck flasks containing medium 88 and incubated for 3-4 days at 80°C in liquid shakers until virus titers reached detectable levels. These 1 L cultures were then used to inoculate 100 L batches of early log *S. solfataricus* P2 grown in the 100 L fermentor (Ace Glass Inc. www.aceglass.com) for 2-3 days at 80°C, pH 3.2. If viral titers reached acceptable levels after 3-5 days (monitored with the TEM) then the 100 L batch was allowed to cool to 22°C and processed. Processing began by draining the fermentor slowly through an in-line

centrifuge (Carr Powerfuge Pilot, Pneumatic Scale Corporation) spinning at 6,000 x g to remove a major portion of the host cells and larger particulate. Filtrate was collected from the in-line centrifuge in 20 L carboys. This large volume of filtrate (still ~100 L) was processed with a tangential flow filter in the same manner as described for filtering and concentration of virus particles described in Appendix A. Briefly a peristaltic pump (Millipore Masterflex I/D, Millipore Corporation) was used to run filtrate out of the 20 L carboys through a S-10 Millipore spiral cartridge tangential flow filter (Millipore Corporation). The volume of the flow through, which is again the filtrate (waste in this case), is reduced to about half of the inflow leaving the retenate (portion containing the virus) at about half of the original volume. Retenate is collected again in 20 L carboys and moved back to the front end of the process to be run through the tangential flow filter again thereby reducing the volume of the retenate by half for each pass of the process. This step is repeated until the original 100 L is concentrated to a volume between 1.0-0.5 L, and then further processed through the Amicon 100 membranes (Millipore Corporation) producing a final volume of 5-8 ml. This concentrate is then loaded on to the same Cs_2SO_4 (38% wt/vol in 25 mM citrate buffer pH 3.5) and centrifuged at 38,000 x g for 20-24 hours. Viral bands are then pulled, dialyzed at 4°C against 25 mM citrate buffer (pH 3.5), and checked for concentration and purity with the TEM. If further concentration is required after this step due to increase in volume from dialysis, it is preformed with the small Microcon 100 filters (Millipore Corporation).

Figure 3.2 - Image of the glass 100 L fermentor (shown with heating element wrapped around) on the left, and the centrifuge filter on the right. A valve is opened at the bottom of the fermentor to run the virus producing culture through the in-line centrifuge when it is ready for processing.

Isolation, Cloning, and Sequencing the Viral Genome

Nucleic acid was extracted from purified virus by standard phenol, chloroform, and Proteinase K plus SDS method (Sambrook et al. 1989). It was then treated with DNase, RNase, Mung Bean, and Pst1 to determine if the genome was DNA or RNA, single strand or double strand in nature. After the nature of the genome was determined, separate restriction endonuclease (RE) libraries were constructed by digestion of the extracted DNA with *Eco*RI, *Pst*I, *Hin*DIII, ligated into BlueScript plasmid vector (Stratagene), and transformed into ultracompetent XL2-Blue cells (Stratagene) using the manufacturer's protocols (http://www.stratagene.com/manuals). A shotgun library was created by Lucigen Corporation. This library differed from the RE libraries in that it was created by randomly shearing the viral DNA and preferentially amplifying pieces that

range in size from 400 bp to 1500 bp with a proprietary method that is successful at amplifying DNA that is hard to clone.

Plasmids were purified from individual colonies from each library and used as templates for Big Dye 3.0 chemistry sequencing reactions as described by the manufacturer (Applied Biosystems). To facilitate sequencing of the RE libraries, sets of primers were constructed to walk along the sequence in from both ends of larger bands (greater than 800 bp), and also to sequence out in order to join a band with its proper neighbor. Walk primers were used to sequence restriction bands greater than 800 bp because the quality of sequence read degrades after approximately 400 bp. It is necessary to create enough good sequence from each direction that it overlaps. New primers were constructed near the ends of each previous run at favorable sites then ~400 bp of new sequence from those new sites towards the middle is gained. This process is repeated until the reads from both ends extend past each other in the middle of that band. The "gap" primers were constructed after the end sequence of the bands was determined. This was done by selecting primer sites in the newly determined sequence to verify its position relative to it's RE library neighbors by sequencing in the opposite direction off its end and on to the adjoining piece.

DNA sequencing was performed on an ABI 3700 automated DNA sequencer according to manufacturer's protocols. (Applied BioSystems). Sequence assembly was performed with the program Sequencer version 4.1.2 (Gene Codes), and putative ORFs were identified using NCBI ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Sequence alignments were performed with BLASTP at NCBI

(http://www.ncbi.nlm.nih.gov/BLAST/), TBLASTN, and BLASTN in Biology WorkBench (http://seqtool.sdsc.edu/CGI/BW.cgi) or http://workbench.sdsc.edu). Predicted viral ORFs were compared in the codon usage database (http://www.kazusa.or.jp/codon/) to those of *Sulfolobus* for codon bias.

Infectivity Assays

Infectivity assays were carried out by inoculating early log cultures of *S. solfataricus* strain P1 (Zillig et al. 1980) and *S. solfataricus* P2 (She et al. 2001) as a replicate with 20 µl of purified virus. Purified virus was treated separately with either 10% Proteinase K (Ambion) heated to 55°C for 30 minutes, 25% chloroform at 22°C for five minutes, 0.3% TritonX 100 at 22°C for three minutes, or with 0.1% SDS at 22°C for three minutes. Control cultures were inoculated with untreated purified virus. All infectivity assays were monitored every 24 hours for virus accumulation by TEM. Virion stability was monitored by TEM visualization of purified virus that was treated in a pH range from 3-6. The pI of the virus was determined by measurement with the PALS Zeta Potential Analyzer (Brookhaven Instrument Corp).

Identification of Major Virion Protein: (performed in Cooperation with Eric Gillitzer, Postdoctoral Fellow in the Young Lab)

Purified virus was denatured and separated onto 12% SDS-PAGE gels, stained with Coomassie Brilliant Blue R250 (Sigma) and the major protein band excised. The gel slice was cut into smaller pieces, de-stained using a 50% acetonitrile in 25 mM ammonium bicarbonate pH 7.9, and desiccated. Ten to fifteen microliters of sequencing grade trypsin (15 ng/µL; Promega) in 25 mM ammonium bicarbonate pH 7.9 was added

to the dehydrated gel slices and allowed to incubate overnight at 37°C. The digestion was stopped by the addition of 20 µL of 50% acetonitrile in 25 mM ammonium bicarbonate pH 7.9 with 1% trifluoroacetic acid (TFA). Matrix for MALDI-TOF was prepared by mixing twice-recrystallized α-cyano-4-hydroxycinnamic acid in a 50% acetonitrile, 0.1% TFA. One-half microliter of sample and 0.5 µL of matrix was spotted onto a MALDI-TOF plate and allowed to dry. Spectra were collected using a Bruker Biflex III MALDI-TOF mass spectrophotometer (Bruker Daltonics) and calibrated using the tryptic digest fragments. Results were compared with theoretical trypsin digests of viral ORF's using Peptide Mass at ExPASy (http://us.expasy.org/tools/peptide-mass.html).

CryoTEM and Image Reconstruction: (performed in Cooperation with Liang Tang Postdoctoral Fellow in the Johnson Lab, Scripps Research Institute)

Frozen-hydrated virus samples were prepared on holey EM grids as described (Dubochet et al. 1988). Briefly, 5 µl of purified virus sample concentrated to approximately 0.5 mg/ml was applied for two minutes onto a previously glow-discharged copper grid coated with a holey carbon film. The grid was blotted with preheated Whatman #2 filter paper to near dryness, flash frozen in a slush of liquid ethane, and transferred into liquid nitrogen. The grid was transferred with a Gatan 626 cryostage (Gatan, Inc.) into a Phillips CM120 transmission electron microscope (Philips/FEI) and operated at an accelerating voltage of 100 kV. Focal pairs of electron micrographs were recorded under low-dose conditions (5 to 10 electrons/$Å^2$) at a magnification of 28,000. Micrographs were digitized using a Zeiss SCAI scanner (Z/I Imaging Corporation). The step size for scanning was 7 µm, corresponding to a pixel size of 2.5 Å. Particles were

boxed with the Autobox subroutine in the program EMAN (Ludtke et al. 1999), and image reconstruction was carried out with the program SPIDER (Frank et al. 1996). A total of 718 particles from micrographs recorded at an under focus of 2.3 µm yielded a reconstruction at 34 Å resolution. The resolution was subsequently improved to 27 Å using an additional 612 particles recorded at an under focus of 1.5 µm. No further contrast transfer focus (CTF) correction was performed, as the resolution was lower than the first zero of the CTF. The resolution was estimated with the Fourier shell correlation method using a criterion of 0.5 (Van Heel 1987).

Results

Single colony isolation established a pure virus-producing culture. Phylogenetic analysis of the 16S rRNA gene of isolate YNPRC179 identified the viral host as *S. solfataricus* (Zillig et al. 1980), an aerobic facultative chemolithotroph (Figure 3.3).

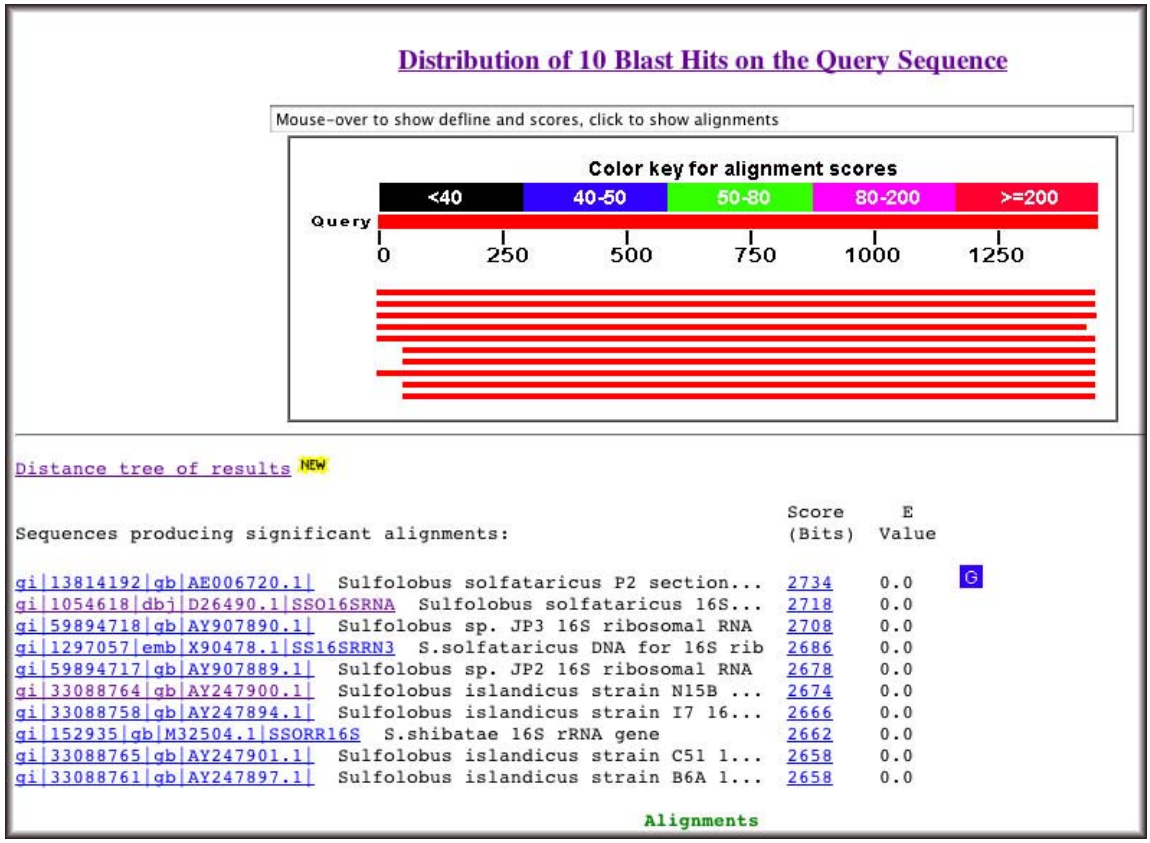

Figure 3.3 - Graphical representation of the of the top ten alignments to the sequence of the 16S rRNA gene of isolate YNPRC179 using Mega BLAST in NCBI's Basic Local Alignment Search Tool (BLAST) site (http://www.ncbi.nlm.nih.gov/BLAST/).

Virus Isolation

Under the optimal growth conditions of 80ºC and pH 3.3 in a minimal media, virus particles were observed directly in culture supernatants by TEM analysis. Virus particles often appeared within 24-48 hours and usually were at the most concentrated levels between day three and five. After four or five days in culture the virus particles often began to look degraded and often were missing their turrets. Normally maximum concentrations would run from five to 20 particles per field of view at a magnification of 20K, although often it would be much lower requiring the viewing of many fields to find

just a few particles. Virus numbers varied quite a bit from batch to batch, and we could not correlate titers with variables such as alternate medias or slight differences in pH of the same media. Alternate hosts were tried also, such as *S. solfataricus* P1 and P2 to maximize production, but we could not get consistent enough results to say whether one host or media was really better than the rest overall. A couple of times we did achieve virus concentrations that approached 0.5 mg/ml (Figure 3.5), but mostly production ran well below that.

The virus was purified to near homogeneity after sequential filtration of culture supernatants and isopycnic banding on Cs_2SO_4 gradients. Virus particles banded at 1.236 g/cm^3 in Cs_2SO_4 and maintained the same morphology as originally observed in the primary enrichment cultures (Figure 3.5).

Infectivity

Purified virus retained its infectivity as demonstrated by its ability to multiply when introduced into virus-free isolates of *S. solfataricus* strain P1 and *S. solfataricus* strain P2. The highest amount of virus was observed in *S. solfataricus* P2 after treatment of the purified virus with Proteinase K (approximately 50% more virus than control). The purified virus retained infectivity after treatment with 0.3% Triton X 100 (similar to control). Infectivity was lost after treatment with 0.1% SDS or 25% chloroform (Figures 3.4 and 3.5).

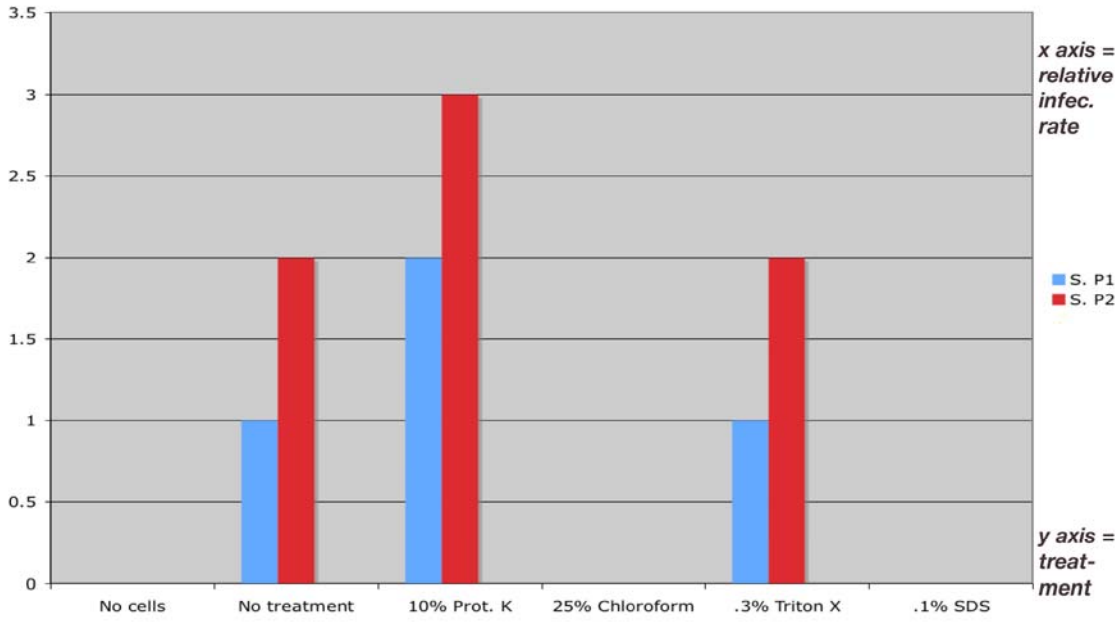

Figure 3.4 – Bar chart showing the infectivity of *S. solfataricus* P1 and P2 with 18 µl of purified virus after various treatments. Rows one and two are the controls. "No cell" signifies that there was no Sulfolobus in the culture and virus added by itself. "No treatment" denotes the presence of P1 and P2 inoculated with virus that had not been treated in any manner.

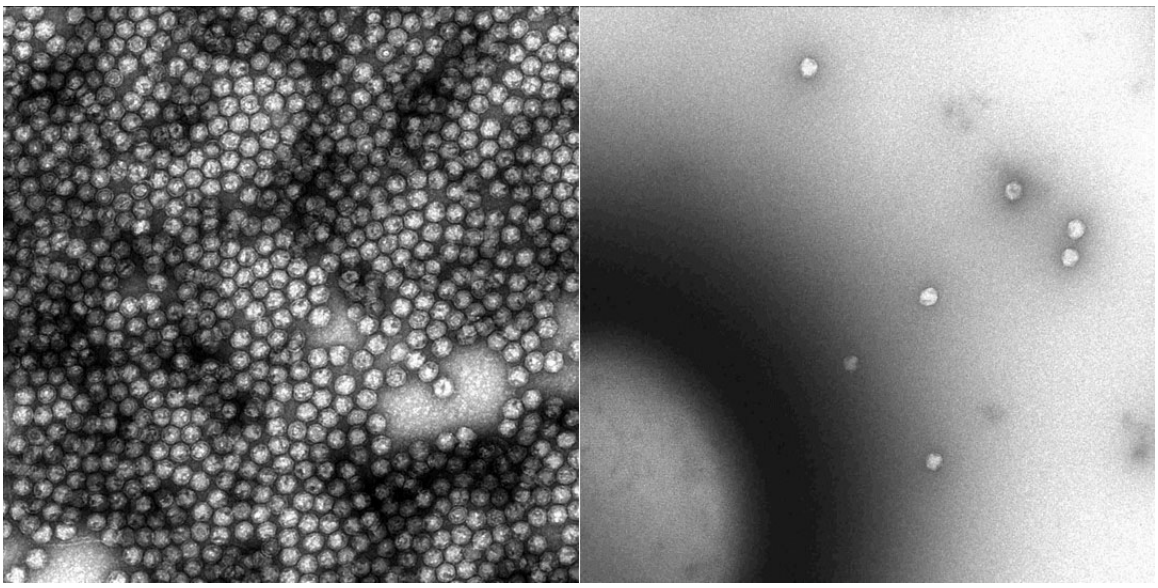

Figure 3.5 - Above left: Image shows a TEM of STIV purified and concentrated from a100 L fermentor batch. Above right: Image shows the host *S. solfataricus* P2 after treatment of the purified virus with Proteinase K.

Southern analysis provided no evidence that the virus integrates its viral genome into the host cell chromosome or induces cell lysis.

Genome Analysis: In cooperation with -
Frank Roberto at Idaho National Lab, and Lucigen Corp.

The genome was initially estimated, from restriction endonuclease digests, to be approximately 19.5 Kbp in length (Figure 3.6) and its treatment with DNase and RNase showed it to be dsDNA (Figure 3.7).

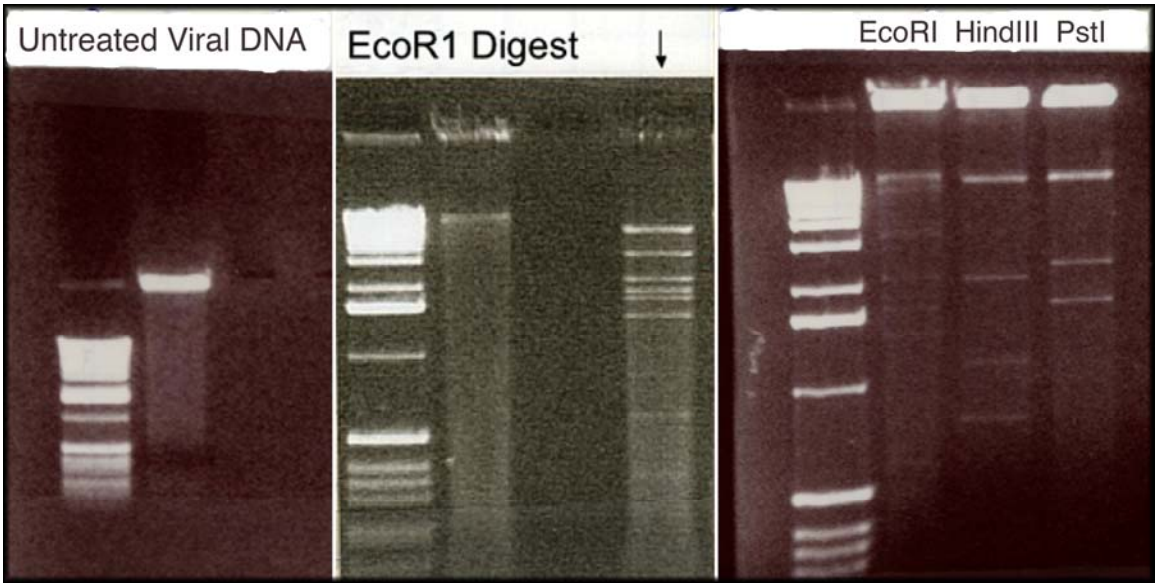

Figure 3.6 - Panel above shows a series of agarose gels stained with ethedium bromide. DNA marker ladder is in the left lane of all of the gels, and the other lanes are loaded with viral DNA treated with the enzyme that is indicated above the lane. The gel on the left is of viral DNA that has not been treated in any manner. This is also the case in the first lane next to the marker in the middle gel. In both controls it is evident that a sizable portion of the DNA stays in the loading well and does not enter the gel. This is also evident in the lanes containing the three restriction digests in the gel on the left.

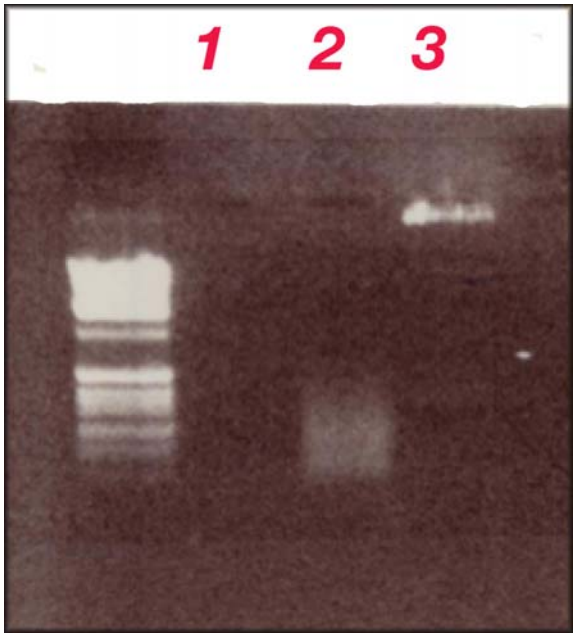

Figure 3.7. Prior to restriction digest library construction, viral DNA was treated with DNase and RNase to determine the nature of the genome. The gel above shows the viral DNA treated with (1) DNase, (2) Mung Bean (single strand cutter), and (3) Pst1.

Constructing multiple RE and shotgun libraries facilitated the sequencing of the viral genome. The restriction libraries were constructed with EcoRI HindIII and PstI as they produced a nice restriction pattern with the viral genome (Figure 3.6). Adding the size of the EcoRI RE library bands together formed a rough genome size estimate of 19.5 Kbp. Primer walks were performed along all of the RE bands we were able to clone until the sequence was joined from either end. Simultaneously gap sequencing was performed off the ends of these bands in order to join them in their proper order. Creating contigs from overlapping sequence of the same libraries as well as crossover from the separate libraries allowed us to build a contiguous sequence of a little over 15 Kbp with a coverage that was often more than 15 x coverage for most bases. We employed Lucigen Corp. to create a shotgun library of the genome as we were not able to clone and

sequence all of the pieces from the RE libraries. With this additional source we were able to close any remaining gaps and join the sequence into a double strand circular genome with a greater than three deep coverage in all areas. The circular nature of the genome was verified by using walk and gap primers already created in the proper order to sequence over the sections that were previously the ends of the sequence before they joined into one contiguous (circular) piece.

The virus encapsulates a 17,663 bp circular dsDNA genome (Figure 3.8). Like its host genome, the viral genome has a low (36%) G+C content (She et al. 2001). A total of 50 ORFs greater than 40 amino acids can be identified. However, this number of predicted ORFs is reduced to 36 predicted ORFs if one requires each ORF to be greater than 50 amino acids and to be associated with TATA-like elements. These 36 ORFs potentially encode for proteins ranging in size from 5.1-57 kDa, and the inferred codon usage of these ORFs shows a codon usage bias that is similar to that of sequenced *S. solfataricus* (Kawarabayasi et al. 2001a; She et al. 2001). ORFs were labeled A B and C for the first three reading frames in the positive direction, and D E and F for the three reading frames on the complementary strand. An overwhelming number of the ORFs that meet all of the above-mentioned criteria are in the positive labeled reading frames (Figure 3.8).

A search of the genome for TATA-like elements associated with the *Sulfolobales* (TTTTTAAA; Soppa 1999) found three exact matches and 16 other TATA-like elements similar to the *Sulfolobales* consensus sequence. If the position of these TATA-like elements is considered along with the degree of overlap between downstream ORFs, it

suggests that 31 of the ORFs are transcribed into six polycistronic RNAs. The other ORFs containing TATA-like elements suggest that they may be transcribed separately.

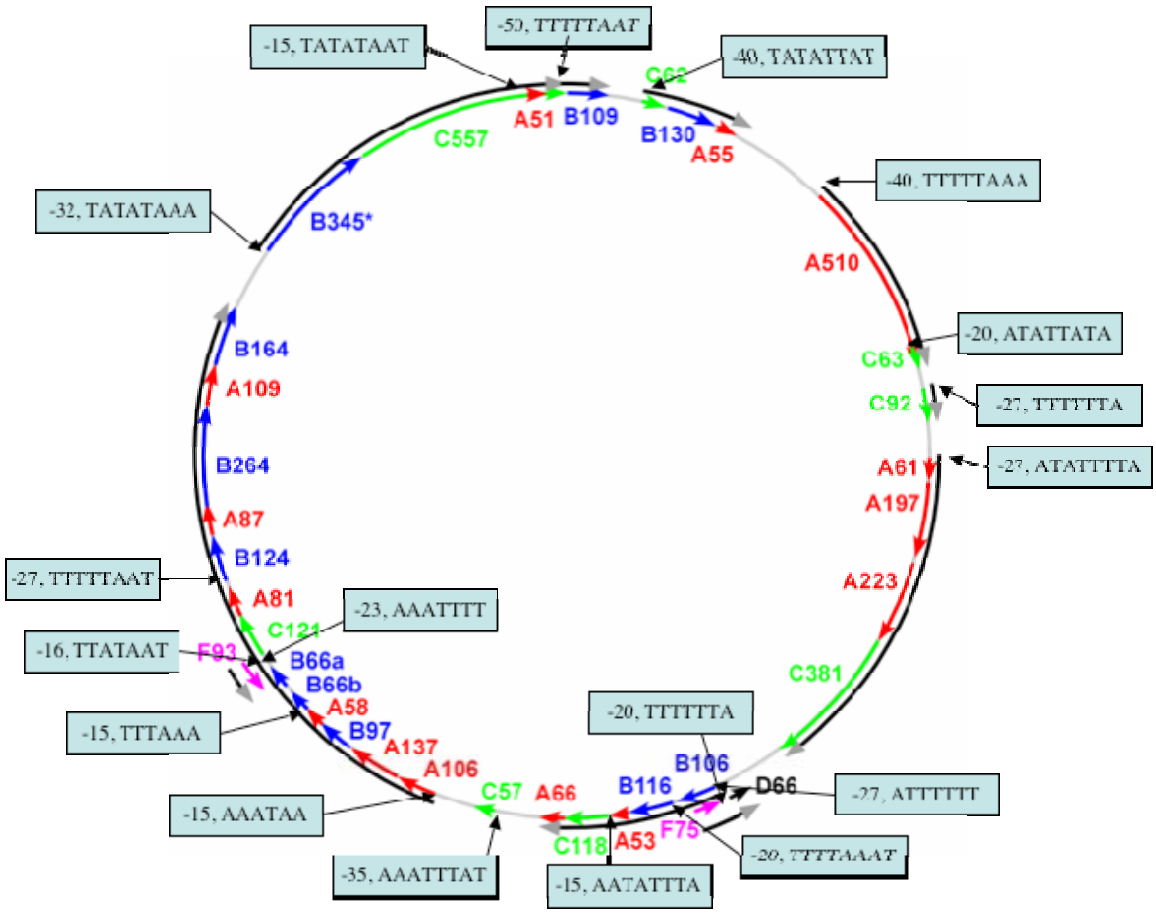

Figure 3.8 - Genome map of the 17633 bp circular dsDNA STIV genome. The 36 predicted open reading frames greater than 50 amino acids in length and associated with TATA-like elements are indicated. Reading frames plus 1 (red), 2 (blue), 3 (green) and reading frames minus 1 (black) and minus 3 (purple) are indicated. The major structural protein (B345) identified by mass spectrometry is indicated with an asterisk. The boxes with arrows around the circular genome show the relative positions downstream of the TATA-like elements and the exact sequence of those elements.

Nearly all of the predicted ORF products show no significant similarity to proteins in the public databases. The exceptions to this observation are a region of ORF C557, which is similar to a hypothetical protein of unknown function encoded by *S. tokodaii*

(Kawarabayasi et al. 2001a), and ORFs B116 and C92 that show similarity to ORFs of the *Sulfolobus* Rudiviruses, SIRV1 and SIRV2 (Prangishvili et al. 1999b). Interestingly, one of these ORFs (ORF 116) is also homologous to an ORF found in the Lipothrixviruses SIFV (Arnold et al. 2000b), and AFV (Bettstetter et al. 2003). They are not clustered in the genome of any of these viruses.

Table 3.2 - Predicted ORFs in the STIV genome that are most similar to other putative ORFs in the database. The similarity scores, along with the organism that the comparative ORF is in, are listed along with which database the sequence was found in, and the type of blast search that was conducted in NCBI to obtain these results.

ORF	Score	HIT	Data Base	Search Type	Match Identity
C557	5.00E-43	Sulfolobus tokodaii complete seq	GBBCT	TBLASTN	Related Host Hypothetical Protein
B345	6.00E-01	HIV type 2 partial env gene	GBVRL	BLASTN	Identified as Coat Protein
B164	1.70E-02	PBCV1	GBVRL	TBLASTN	Virus with Internal Lipid Membrane
B116	4.00E-32	SIRV-1	GBVRL	TBLASTN	Virus with same host
B116	2.00E-31	SIRV-2	GBVRL	TBLASTN	Virus with same host
B116	7.00E-08	SIFV	GBVRL	TBLASTN	Virus with same host
A109	1.00E-05	Sulfolobus sulfataricus 166 of 272	GBBCT	BLASTN	Related Host Hypothetical Protein
C92	9.00E-25	SIRV-1	GBVRL	TBLASTN	Virus with same host
C92	5.00E-23	SIRV-2	GBVRL	TBLASTN	Virus with same host

Identification of the Major Structural Protein

SDS-PAGE analysis of purified virions revealed an abundant protein with an estimated molecular mass of 37 kDa, as well as several minor proteins with estimated masses of 25 kDa, 12.5 kDa and 10 kDa (Figure 3.4). The gene encoding the major 37 kDa virion protein was identified by trypsin digestion of the coat protein from purified virus, followed by MALDI mass spectrometry of the product. Five of these fragments matched the theoretical trypsin digests of ORF B345 predicted by ExPASy's Peptide Mass program (Table 3.2). The pI of the virus was determined to be pH 3.26 by dynamic

light scattering zeta potential analysis (Figure 3.9). Other protein bands were not retrieved in sufficient quantities to perform mass spectrometry.

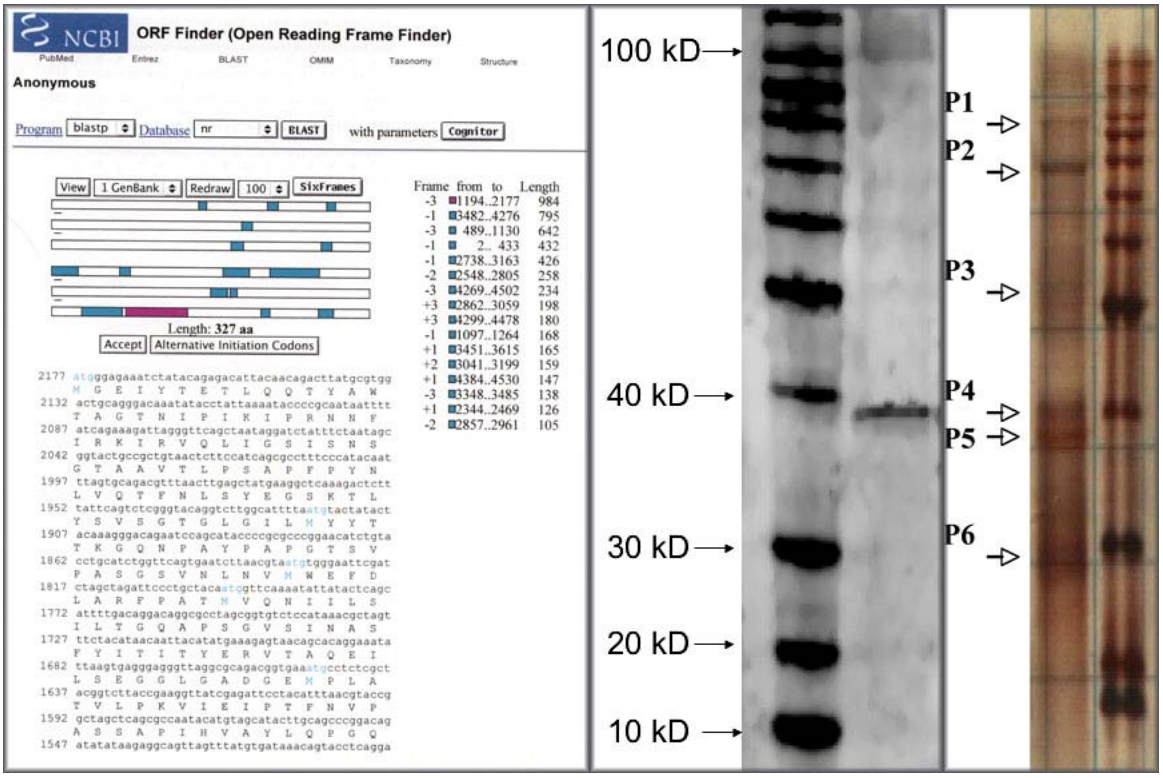

Figure 3.9 - Above left: Image shows the larger ORF's found in NCBI's ORF Finder program. The ORF highlighted in red was subsequently named B345 and was shown to be the ORF coding for the 37 kDd protein identified in the gel (shown in the middle). Putting the sequence making up this ORF into in ExPASy's Peptide Mass program to perform a theoretical trypsin digest of the protein that is coded for by this ORF allowed us to link this ORF to the 37 kDa protein. Middle: Image shows an SDS-PAGE gel with the viral protein band of approximately 37 kDa in the right lane next to the marker land with bands delineated from 15-40 kDa. Gel on the right shows many minor viral proteins run next to a marker. Six of the more prominent bands are marked with P5 being the 37 kDa. P3 is 25 kDa, P2 is 12.5 kDa, and P1 is 10 kDa. P4 may be a modified protein normally seen at 37 kDa.

Table 3.3 - This table shows the largest fragments from the theoretical Trypsin digest of the isolated 37 kDa protein performed in ExPASy's Peptide Mass program. The six fragments marked with checks matched the masses of fragments detected with the MALDI Mass Spectrometer when an actual trypsin digest was carried out with the isolated protein.

mass	position	#MC	peptide sequence
5301.4599	279-323	0	YFNGDLDLTHAPSDSIEYDL ALQNQDNVYSLYVSYVLPYY DHSQR
3971.0436	36-73	0	VQLIGSISNSGTAAVTLPSA PFPYNLVQTFNLSYEGSK
3916.0564	124-159	0	FPATMVQNIILSILTGQAPS GVSINASFYITITYER
✓ 3245.5684	93-123	0	GQNPAYPAPGTSVPASGSVN LNVMWEFDLAR
✓ 3155.5432	249-277	0	VSWAALQAENQAEYQVAPYS GASAIIDFR
✓ 3052.6505	186-213	0	VIEIPTFNVPASSAPIHVAY LQPGQIYK
2729.3491	1-24	0	MGEIYTETLQQTYAWTAGTN IPIK
✓ 2598.2933	215-237	0	QLVYVINSTSGINNTDPTEY ELK
✓ 2596.3538	160-185	0	VTAQEILSEGGLGADGEMPL ATVLPK
✓ 2068.0671	74-92	0	TLYSVSGTGLGILMYYTTK
663.3573	28-32	0	NNFIR
616.3300	241-246	0	GVPTDK
536.3442	324-327	0	YLLK

Batch: 0

Sample ID	**Pure 2-11-03 (Combined)**
Operator ID	**GIII**
Notes	**in 25mM citrate**

Measurement Parameters:

Mean Zeta Potential	= -1.72 mv	Liquid	= Aqueous
Zeta Potential Model	= Smoluchowski	Temperature	= 25.0 °C
Mean Mobility	= -0.13 (μ/s) / (V/cm)	Viscosity	= 0.890 cP
pH	= 3.20	Refractive Index	= 1.330
Conductance	= 2327 μS	Dielectric Constant	= 78.54
Concentration	= 0.10 mg/mL	Particle Size	= 90.0 nm

Instrument Parameters:

Sample Count Rate	= 278 kcps	Voltage	= 4.00 volts
Ref. Count Rate	= 1676 kcps	Electric Field	= 7.38 V/cm
Wavelength	= 661.0 nm	User1	= 0.00
Field Frequency	= 2.00 Hz	User2	= 0.00

Figure 3.10 - Analysis of a pure STIV sample with the PALS Zeta Potential Analyzer (Brookhaven Instruments Corp.) showing, among other things, the pI of the virus and the mean diameter of the particle.

CryoTEM and Image Reconstruction: (In cooperation with – Liang Tang Postdoctoral Fellow in the Johnson Lab)

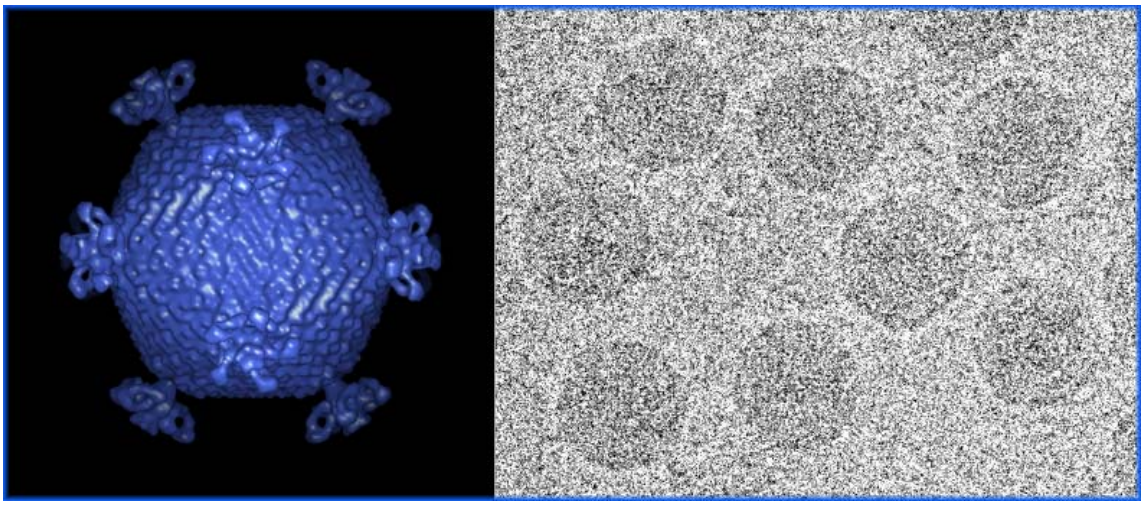

Figure 3.11 - Cryo-transmission electron microscopy and image reconstruction of STIV (above left). Electron micrograph of vitrified STIV sample recorded at an under focus of 1.5 μm (above right).

The image reconstruction at ~27 Å resolution revealed a unique virus architecture that included a complex structure of appendages at each of the 12 five-fold vertices. These appendages are five-sided, turret-like structures that have an average diameter of 24 nm and extend 13 nm above the particle surface. The center of each turret contains an approximate 3 nm channel, which could provide access between the interior and exterior of the virion. There appears to be density at each end of the channel, indicating that the channel may be blocked and suggesting a possible regulatory role associated with this density.

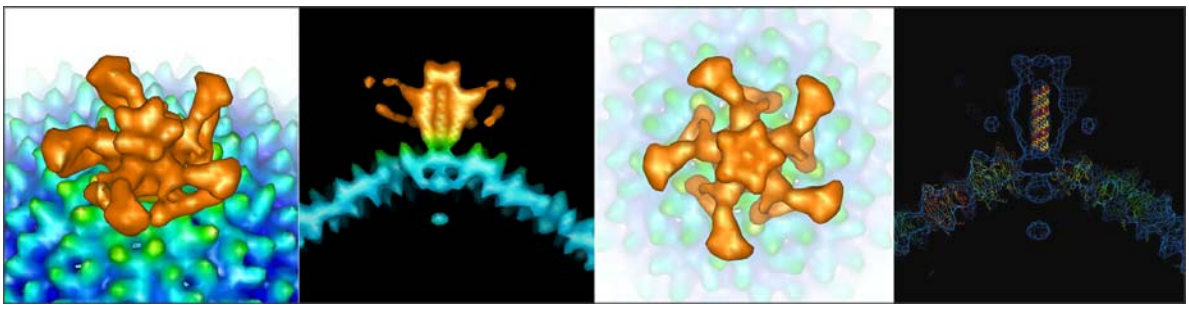

Figure 3.12 - From left to right; Angle cross-section and top view of turret-like projections extending 13 nm above the surface of the virus surface. Cross section showing a 3 nm central channel in a single turret-like projection.

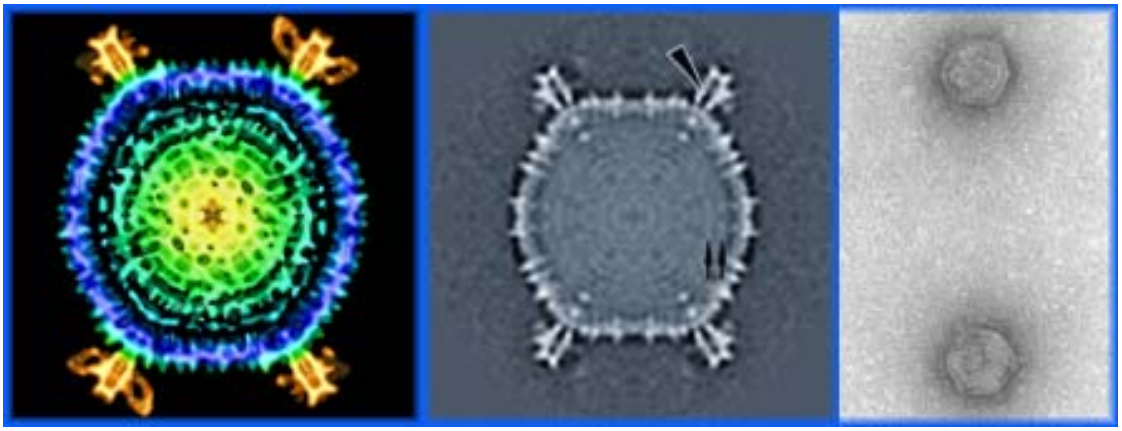

Figure 3.13 - Left: Image shows a 7.5 Å thick equatorial section of the cryoEM map viewed down the icosahedral two-fold axis. Center: An arrowhead and a pair of arrows indicate the channel in the five-fold turret and two leaflets of the putative bilayer, respectively. The image on left is the same electron density map enhanced with color. Right: TEM image taken at 25,000x that alludes to the possibility of a structural bilayer in the virus.

The structural features of the capsid indicate that the virion is built upon a pseudo T=31 icosahedral lattice, which is the first of this type to be observed (Figure 3.14). The diameter of the capsid is ~74 nm with a ~6.4 nm shell thickness. The icosahedral asymmetric unit consists of five trimers (pseudo hexamers) of the major capsid protein plus one additional minor capsid protein at the five-fold vertex.

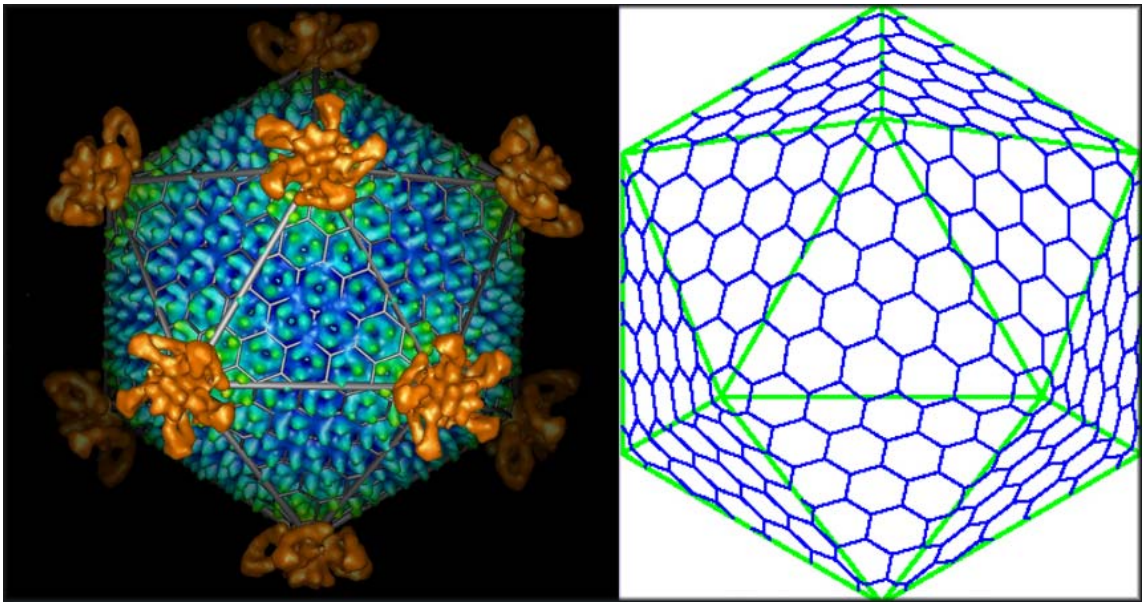

Figure 3.14 - Left: STIV surface features with turret-like projections extending from each of the five-fold vertices, overlaid with the surface capsomer architecture of T=31. Right: Rendering of the virus stripped of surface features and accentuating its icosahedral nature and the T=31 symmetry unique to STIV.

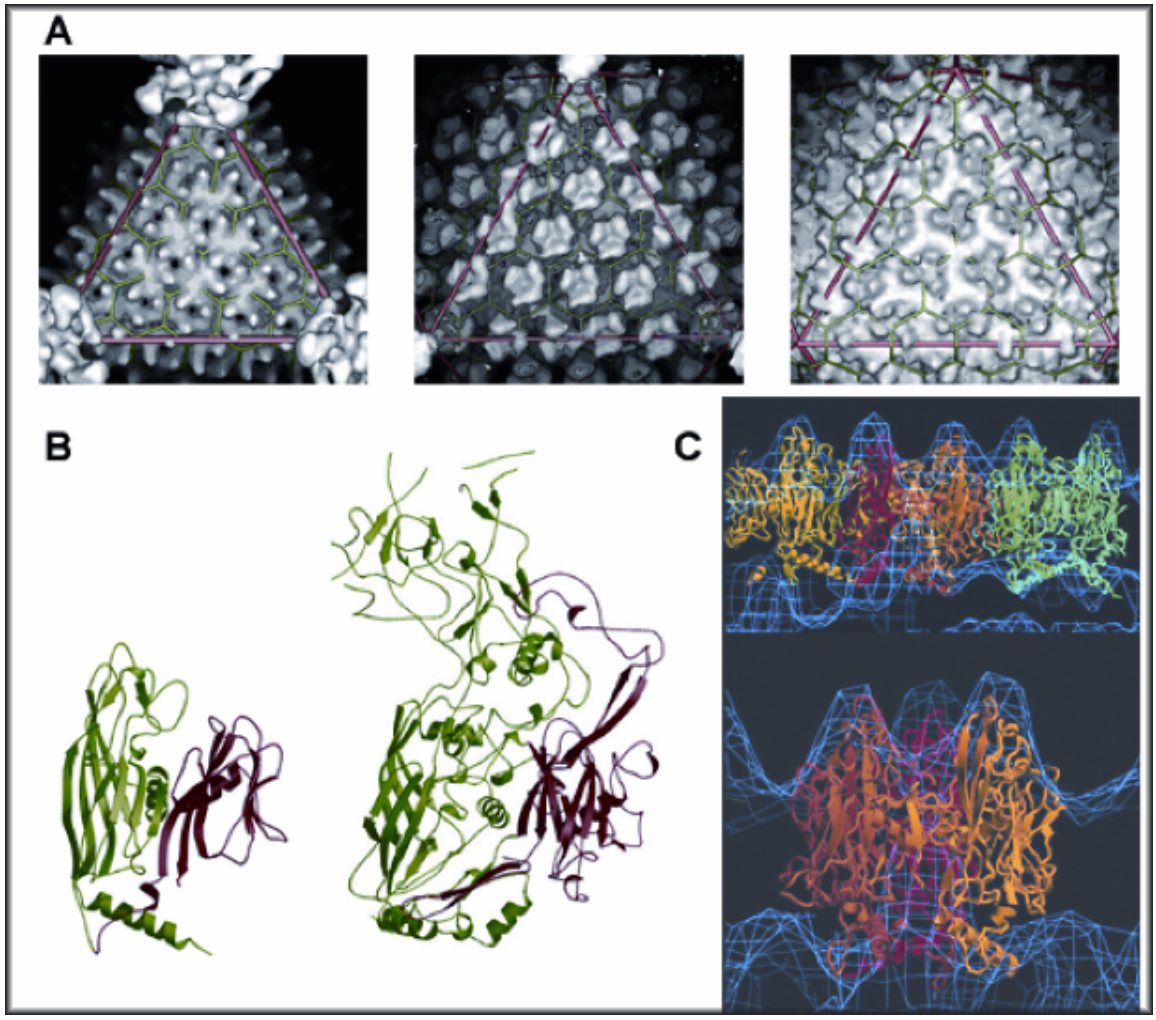

Figure 3.15 - (A) Surface capsomer architecture of T=31 STIV (left), adenovirus (center) and PRD1 (right) The icosahedral asymmetric unit (triangle) consists of five trimers (pseudo-hexamers) of the major capsid protein plus one additional minor capsid protein at the five-fold vertex. (B) The X-ray crystal structures (ribbon diagrams) of PRD1 (29) (left) and adenovirus hexon (29) (right) major coat proteins showing their similar folds. Both contain the "double-jellyroll" motif (red and green). (C) Fitting of the PRD1 P3 coat protein X-ray crystal structure into the STIV cryo-image reconstruction density. The arrangements of the two beta-sandwich domains in STIV are remarkably similar to those in PRD1 and adenovirus, as indicated by the characteristic location of the towers on the outer surface of the major capsid protein with respect to the quasi-equivalent lattice in each case.

Discussion

In this chapter we reported on the first structural and genetic characterization of an icosahedral virus from a hyperthermo-acidophilic archaeon. The virus was originally isolated from environmental samples taken from hot springs in the Rabbit Creek thermal area located in the Midway Geyser Basin of YNP. The Rabbit Creek springs are characterized by high temperature (70-90°C), acidic pH (2-4), and water geochemistry that is dominated by high levels of sulfate, silicon, iron, sodium and aluminum (Braden Hanna, *unpublished data*). We have also detected similar virus particle morphologies from other high temperature acidic thermal features in YNP (Mud Volcano area), suggesting that this virus occurs commonly in YNP.

It is surprising that we are unable to detect a significant similarity between the potential gene products encoded on the viral genome with those available in the public databases. A similar lack of homology to sequences in the public databases has also been observed in the genome analysis in the sequenced viral genomes of Fuselloviruses *Sulfolobus* spindle-shaped virus SSV; (Palm et al. 1991; Stedman et al. 2003), Rudiviruses *S. icelandicus* rod-shaped virus SIRV; (Prangishvili et al. 1999b), and Lipothrixviruses SIFV (Arnold et al. 2000b) and AFV1 (Bettstetter et al. 2003) that also replicate in *Sulfolobales* hosts. This suggests that viruses of the *Sulfolobales* may have unique processes for carrying out their replication cycles and/or may reflect unique requirements for replication in hyper-thermophilic archaeal hosts. Perhaps the ORF conserved in all except the Fuselloviruses is involved in this replication or adaptation to

extreme conditions. It is also possible that these viruses can undergo rapid evolution, or that the unique environment they inhabit has kept them ecologically isolated for a long period of time, obscuring their relationships at the gene level to other viruses and organisms. It seems probable that a combination of these factors contribute to the genetic diversity of these viruses, but the relative ecological isolation of the *Sulfolobales* and their viruses may be the most important factor that has resulted in their unique genome composition (Whitaker et al. 2003).

The elaborate turret-like structures evident in the reconstruction of the viral capsid are the predominant surface features. It is possible that these structures may play a role in host recognition and/or attachment. We speculate that the central channel in each turret might act as a portal for translocation of the viral nucleic acid as is typical of the five-fold vertices in many viruses (Bothner et al. 1998; Lewis et al. 1998).

Ultimately, analysis of the structural features of this virus allowed us speculate about its relationship to other Archaeal thermal viruses as well possible structural similarities to viruses in both of the other domains of life, and the possible evolutionary ramifications that might be inferred from these similarities. The structural organization of the STIV capsid is similar to other viruses with large T numbers, such as human adenovirus (Stewart et al. 1991), bacteriophage PRD1 (Bamford et al. 1995; Butcher et al. 1995) and algal virus PBCV-1 (Nandhagopal et al. 2002). The predicted secondary structure of the STIV major capsid protein (see above) is predominately β-sheet. This is similar to the secondary structure of the major capsid proteins of adenovirus, PRD1, and PBCV-1 as determined by X-ray crystallography. The atomic structures of these three

capsid proteins show a common two β-sandwich fold (Figure 3.15b). To investigate the relationship between these capsid proteins, the atomic coordinates of the major capsid protein from PRD1 and Adenovirus were docked into the three-dimensional map of STIV (Figure 3.15c). These docking studies suggest that STIV also has an arrangement of two β-sandwich domains that are remarkably similar to those in PRD1 and adenovirus. This is indicated by the characteristic location of the turrets on the outer surface of the major capsid protein with respect to the quasi-equivalent lattice in each case. All three of these viruses also employ a pseudo-hexameric structure composed of six β-sandwich domains as the basic building block to fulfill the quasi-equivalent assembly of the capsid. Like PRD1 and PBCV-1, the cryo EM map of STIV suggests that it may contain an internal lipid envelope. Biochemical and biophysical analyses are underway to confirm the presence of an inner lipid envelope.

The lack of genomic similarity between these viruses necessitates looking at other similarities to infer relationships. An alternative approach for comparison among long-separated viral lineages is the possibility that they share a common innate “self” based on the structural and assembly principles of the virion. The proposal that molecular organization (folding) of biological macromolecules preceded the appearance of early life forms suggests that there may be a conservation of structure (and function) that may be useful for penetrating deep into the history of life (Bamford 2003). The similar organization of the capsid proteins from the bacteriophage PRD1, the eukaryotic adenovirus, and now the archaeal STIV suggests long-range evolutionary relationships between these viruses from all three domains of life.

There are other potential explanations for this astounding similarity. First, there could have been convergent evolution, although this is unlikely as there are many different virus morphologies that can infect Archaea, Eukarya, and Bacteria. A second explanation is that there was horizontal gene transfer between all three domains, a possibility that cannot be ruled out but is also unprecedented and must have happened long ago for all sequence resemblance to have disappeared. Finally, a common ancestor responsible for the innate self still evident in STIV, PRD1, PBCV-1 and adenovirus has been postulated, and would have preceded the divergence of the three domains of life over three billion years ago. We believe that the striking similarity in the characteristics of these three major virion proteins of these viruses argues for the latter.

The work reported here further supports the concept that the viral world has lineages that can be traced back to near the root of the universal tree of life (Bamford 2003). Further characterizations of STIV and other viruses from hyperthermophilic archaea are likely to greatly expand our knowledge of evolution of life on earth and provide new insights into the archaeal domain of life.

CHAPTER 4

CONCLUSIONS TO QUESTIONS AND HYPOTHESIS

Question 1

Our first major question addressed by this study was, are there detectable viruses present in the thermal features of YNP? This was the first investigation to address this question in YNP, and we definitively determined that there are thermal viruses present in these environments. We obtained direct visual evidence with the TEM, as well as supporting genomic evidence, that there are numerous viruses and VLPs present both in number and morphotype in these environments.

Question 2

In determining which methodology would be most effective for isolating thermal virus from YNP, we found that it was more productive to use a host dependant approach. We had problems developing any sort of yield from the host independent method of filter concentrating large quantities of sample to detect virus. Culturing host organisms from the environment, and then isolating virus from these cultures proved to be an effective method for performing a survey of viruses present in the thermal acidic features of YNP.

Question 3

In answering the question of whether we could isolate thermal viruses from YNP, and if we could how similar would they be to other thermal viruses previously characterized, we found that we could isolate several new viruses and that some of them

were morphologically similar to those that had been seen before and that many of them were quite unique. Furthermore we found that not only were there several new morphotypes present in these environment, but that the viruses that appeared similar to those characterized previously were substantially different on a genomic level. This genomic diversity meant that not only did their genes not match well to there presumed thermal cousins, but they contained little to no matches to anything else in the database.

Question 4

The second portion of this study was focused upon answering the question of whether a new thermal virus could be characterized through this work, and what that might tell us about the relationships of thermal viruses to one another and to other viruses in general. Through this effort (by isolating and characterizing STIV) we were able to provide the first ever cryo-EM structure of a thermal virus, describe the first icosahedral thermal virus, elucidate a unique virion symmetry (T = 31), discover the first thermal virus with an internal lipid bi-layer, and provide structural evidence for a likely connection between entities from all three domains of life. The cryo-EM structural characterization of STIV allowed us to speculate about the significance of the commonality of a fold in its major coat protein that it shares with PRD1 and Adenovirus—two other viruses that infect hosts residing in the two other domains of life. This shared feature suggests the possibility that they all hail from a common ancestor that predates the divergence of the three separate domains of life.

The analysis of STIV's genome produced no major surprises in so far as the ORFs held little similarity to other characterized thermal Archaeal virus and almost none to any other genes in the database period.

Hypothesis

We hypothesize that there are viruses present in the thermal acidic environments of YNP and that the isolation and characterization of these viruses will increase the understanding of thermal virology in general. Through this research we determined that there are indeed viruses present in the thermal acidic environments of YNP. Furthermore it was evident that these amazing viruses are not only plentiful in number but also varied in nature, offering untold future research opportunities both in virology and the biochemistry of extreme thermal environments. Analysis of structural and genomic data obtained in this study did add to the limited knowledge of relationships between thermal viruses themselves, and allowed us to speculate about relationships of theses viruses to viruses belonging to the other two domains of life. This being said, it is obvious that with this research we felt as if we have barely scratched the surface of what might be found in these environments.

REFERENCES CITED

Achenbach, L., & Woese, C. (1995). 16s and 23s rrna-like primers. In F. T. Robb & A. R. Place (Eds.), *Thermophiles* (Vol. 3, pp. 201-203). Cold Spring Harbor: Cold Spring Harbor Press.

Achenbach-Richter, L., Gupta, R., Stetter, K. O., & Woese, C. R. (1987). Were the original eubacteria thermophiles? *Syst Appl Microbiol, 9*, 34-39.

Ackermann, H. W. (1987). Bacteriophage taxonomy in 1987. *Microbiol Sci, 4*(7), 214-218.

Ackermann, H. W. (1996). Frequency of morphological phage descriptions in 1995. *Arch Virol, 141*(2), 209-218.

Ackermann, H. W. (2001). Frequency of morphological phage descriptions in the year 2000. Brief review. *Arch Virol, 146*(5), 843-857.

Ackermann, H. W. (2003). Bacteriophage observations and evolution. *Res Microbiol, 154*(4), 245-251.

Ahmed, H., Ettema, T. J., Tjaden, B., Geerling, A. C., van der Oost, J., & Siebers, B. (2005). The semi-phosphorylative entner-doudoroff pathway in hyperthermophilic archaea: A re-evaluation. *Biochem J, 390*(Pt 2), 529-540.

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped blast and psi-blast: A new generation of protein database search programs. *Nucleic Acids Research, 25*, 3389-3402.

Arnold, H.-P., She, Q., Phan, H., Stedman, K., Prangishvili, D., Holz, I., et al. (1999a). The genetic element pssvx of the extremely thermophilic crenarchaeon *sulfolobus* is a hybrid between a plasmid and a virus. *Molecular Microbiology, 34*(2), 217-226.

Arnold, H. P., Stedman, K. M., & Zillig, W. (1999b). Archaeal phages. In A. Granoff & R. G. Webster (Eds.), *Encyclopedia of virology, second edition* (Vol. 1, pp. 76-89). London: Academic Press.

Arnold, H. P., Ziese, U., & Zillig, W. (2000a). Sndv, a novel virus of the extremely thermophilic and acidophilic archaeon sulfolobus. *Virology, 272*(2), 409-416.

Arnold, H. P., Zillig, W., Ziese, U., Holz, I., Crosby, M., Utterback, T., et al. (2000b). A novel lipothrixvirus, sifv, of the extremely thermophilic crenarchaeon sulfolobus. *Virology, 267*(2), 252-266.

Bamford, D. H. (2003). Do viruses form lineages across different domains of life? *Res Microbiol, 154*(4), 231-236.

Bamford, D. H., Caldentey, J., & Bamford, J. K. (1995). Bacteriophage prd1: A broad host range dsdna tectivirus with an internal membrane. *Adv Virus Res, 45*, 281-319.

Barns, S. M., Delwiche, C. F., Palmer, J. D., & Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily, and monophyly from environmental rrna sequences. *Proceedings of the National Academy of Science USA, 93*, 9188-9193.

Barns, S. M., Fundyga, R. E., Jeffries, M. W., & Pace, N. R. (1994). Remarkable archaeal diversity detected in a yellowstone national park hot spring environment. *Proceedings of the National Academy of Sciences, 91*, 1609-1613.

Baroudy, B. M., & Moss, B. (1982). Sequence homologies of diverse length tandem repetitions near ends of vaccinia virus genome suggest unequal crossing over. *Nucleic Acids Res, 10*(18), 5673-5679.

Bell, S. D., Cairns, S. S., Robson, R. L., & Jackson, S. P. (1999). Transcriptional regulation of an archaeal operon in vivo and in vitro. *Mol Cell, 4*(6), 971-982.

Bell, S. P., & Dutta, A. (2002). DNA replication in eukaryotic cells. *Annu Rev Biochem, 71*, 333-374.

Benzer, S. (1955). Fine structure of a genetic region in bacteriophage. *Proc Natl Acad Sci U S A, 41*(6), 344-354.

Bettstetter, M., Peng, X., Garrett, R. A., & Prangishvili, D. (2003). Afv1, a novel virus infecting hyperthermophilic archaea of the genus acidianus. *Virology, 315*(1), 68-79.

Blum, H., Zillig, W., Mallok, S., Domdey, H., & Prangishvili, D. (2001). The genome of the archaeal virus sirv1 has features in common with genomes of eukaryal viruses. *Virology, 281*(1), 6-9.

Borsheim, K. Y., Bratbak, G., & Heldal, M. (1990). Enumeration and biomass estimation of planktonic bacteria and viruses by transmission electron microscopy. *Appl Environ Microbiol, 56*(2), 352-356.

Bothner, B., Dong, X. F., Bibbs, L., Johnson, J. E., & Siuzdak, G. (1998). Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J Biol Chem, 273*(2), 673-676.

Brenner, S. (1961). Rna, ribosomes, and protein synthesis. *Cold Spring Harb Symp Quant Biol, 26*, 101-110.

Brierley, C. L., & Brierley, J. A. (1973). A chemoautotrophic and thermophilic microorganism isolated from an acid hot spring. *Can J Microbiol, 19*(2), 183-188.

Brock, T. D. (1995). The road to yellowstone--and beyond. *Annu Rev Microbiol, 49*, 1-28.

Brock, T. D. (1997). The value of basic research: Discovery of thermus aquaticus and other extreme thermophiles. *Genetics, 146*(4), 1207-1210.

Butcher, S. J., Bamford, D. H., & Fuller, S. D. (1995). DNA packaging orders the membrane of bacteriophage prd1. *Embo J, 14*(24), 6078-6086.

Chen, F., Suttle, C. A., & Short, S. M. (1996). Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. *Applied and Environmental Microbiology, 62*(8), 2869-2874.

Cohen, S. G. (1995). Pliny the elder (23-65 a.D.) roman historian and encyclopedist. *Allergy Proc, 16*(4), 212-213.

Dahlke, I., & Thomm, M. (2002). A pyrococcus homolog of the leucine-responsive regulatory protein, lrpa, inhibits transcription by abrogating rna polymerase recruitment. *Nucleic Acids Res, 30*(3), 701-710.

DeLong, E. (2001). {1} a phylogenetic perspective on hyperthermophilic microorganisms. *Academic Press*, 3-11.

Dubochet, J., Adrian, M., Chang, J. J., Homo, J. C., Lepault, J., McDowall, A. W., et al. (1988). Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys, 21*(2), 129-228.

Filee, J., Forterre, P., & Laurent, J. (2003). The role played by viruses in the evolution of their hosts: A view based on informational protein phylogenies. *Res Microbiol, 154*(4), 237-243.

Forterre, P. (2003). The great virus comeback-- from an evolutionary perspective. *Res Microbiol, 154*(4), 223-225.

Forterre, P., Charbonnier, F., Marguet, E., Harper, F., & Henckes, G. (1992). Chromosome structure and DNA topology in extremely thermophilic archaebacteria. *Biochem Soc Symp, 58*, 99-112.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., et al. (1996). Spider and web: Processing and visualization of images in 3d electron microscopy and related fields. *J Struct Biol, 116*(1), 190-199.

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature, 399*(6736), 541-548.

Fuhrman, J. A., McCallum, K., & Davis, A. A. (1993). Phylogenetic diversity of subsurface marine microbial communities from the atlantic and pacific oceans. *Appl Environ Microbiol, 59*(5), 1294-1302.

Garrity, G. M. (2001). *The archaea and the deeply branching and phototrophic bacteria* (2nd ed. Vol. 1). New York: Springer.

Geslin, C., Le Romancer, M., Gaillard, M., Erauso, G., & Prieur, D. (2003). Observation of virus-like particles in high temperature enrichment cultures from deep-sea hydrothermal vents. *Res Microbiol, 154*(4), 303-307.

Haring, M., Peng, X., Brugger, K., Rachel, R., Stetter, K. O., Garrett, R. A., et al. (2004). Morphology and genome organization of the virus psv of the hyperthermophilic archaeal genera pyrobaculum and thermoproteus: A novel virus family, the globuloviridae. *Virology, 323*(2), 233-242.

Haring, M., Vestergaard, G., Rachel, R., Chen, L., Garrett, R. A., & Prangishvili, D. (2005). Virology: Independent virus development outside a host. *Nature, 436*(7054), 1101-1102.

Hendrix, R. W. (1999). Evolution: The long evolutionary reach of viruses. *Curr Biol, 9*(24), R914-917.

Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J Gen Physiol, 36*(1), 39-56.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, Takahashi, Sekine, et al. (2001a). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon *sulfolobus tokodaii* strain 7. *DNA Research, 8*(4), 123-140.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., et al. (2001b). Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, sulfolobus tokodaii strain7. *DNA Res, 8*(4), 123-140.

Kessler, A., Brinkman, A. B., van der Oost, J., & Prangishvili, D. (2004). Transcription of the rod-shaped viruses sirv1 and sirv2 of the hyperthermophilic archaeon sulfolobus. *J Bacteriol, 186*(22), 7745-7753.

Kirk, D. L., & Gruber, H. (2005). Ferdinand cohn, multi-faceted microbiologist extraordinaire. *Protist, 156*(3), 355-358.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene, 234*(2), 187-208.

Kyrpides, N. C., & Ouzounis, C. A. (1999). Transcription in archaea. *Proc Natl Acad Sci U S A, 96*(15), 8545-8550.

Laksanalamai, P., & Robb, F. T. (2004). Small heat shock proteins from extremophiles: A review. *Extremophiles, 8*(1), 1-11.

Lee, S. J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., et al. (2003). Trmb, a sugar-specific transcriptional regulator of the trehalose/maltose abc transporter from the hyperthermophilic archaeon thermococcus litoralis. *J Biol Chem, 278*(2), 983-990.

Lewis, J. K., Bothner, B., Smith, T. J., & Siuzdak, G. (1998). Antiviral agent blocks breathing of the common cold virus. *Proc Natl Acad Sci U S A, 95*(12), 6774-6778.

Ludtke, S. J., Baldwin, P. R., & Chiu, W. (1999). Eman: Semiautomated software for high-resolution single-particle reconstructions. *J Struct Biol, 128*(1), 82-97.

Madigan, M. T., & Marrs, B. L. (1997). Extremophiles. *Sci Am, 276*(4), 82-87.

Nadal, M., Mirambeau, G., Forterre, P., Reiter, W.-D., & Duguet, M. (1986). Positively supercoiled DNA in a virus-like particle of an archaebacterium. *Nature, 321*, 256-258.

Nandhagopal, N., Simpson, A. A., Gurnon, J. R., Yan, X., Baker, T. S., Graves, M. V., et al. (2002). The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. *Proc Natl Acad Sci U S A, 99*(23), 14758-14763.

Nathans, D., & Smith, H. O. (1975). Restriction endonucleases in the analysis and restructuring of dna molecules. *Annu Rev Biochem, 44*, 273-293.

Nisbet, E. G., & Sleep, N. H. (2001). The habitat and nature of early life. *Nature, 409*(6823), 1083-1091.

Noble, R. T., & Fuhrman, J. A. (2000). Rapid virus production and removal as measured with fluorescently labeled viruses as tracers. *Appl Environ Microbiol, 66*(9), 3790-3797.

Noyer-Weidner, M., & Schaffner, W. (1995). Felix hoppe-seyler (1825-1895) a pioneer of biochemistry and molecular biology. *Biol Chem Hoppe Seyler, 376*(8), 447-448.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., & Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc Natl Acad Sci U S A, 59*(2), 598-605.

Ortmann, A. C., Wiedenheft, B., Douglas, T., & Young, M. (2006). Hot crenarchaeal viruses reveal deep evolutionary connections. *Nat Rev Microbiol, 4*(7), 520-528.

Ouhammouch, M., Dewhurst, R. E., Hausner, W., Thomm, M., & Geiduschek, E. P. (2003). Activation of archaeal transcription by recruitment of the tata-binding protein. *Proc Natl Acad Sci U S A, 100*(9), 5097-5102.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science, 276*, 734-740.

Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W.-D., et al. (1991). Complete nucleotide sequence of the virus ssv1 of the archaebacterium *sulfolobus shibatae*. *Virology, 185*(1), 242-250.

Paul, J. H., Jiang, S. C., & Rose, J. B. (1991). Concentration of viruses and dissolved DNA from aquatic environments by vortex flow filtration. *Appl Environ Microbiol, 57*(8), 2197-2204.

Paul, J. H., Rose, J. B., Jiang, S. C., Kellogg, C. A., & Dickson, L. (1993). Distribution of viral abundance in the reef environment of key largo, florida. *Appl Environ Microbiol, 59*(3), 718-724.

Peng, X., Blum, H., She, Q., Mallok, S., Brugger, K., Garrett, R. A., et al. (2001). Sequences and replication of genomes of the archaeal rudiviruses sirv1 and sirv2: Relationships to the archaeal lipothrixvirus sifv and some eukaryal viruses. *Virology, 291*(2), 226-234.

Peng, X., Kessler, A., Phan, H., Garrett, R. A., & Prangishvili, D. (2004). Multiple variants of the archaeal DNA rudivirus sirv1 in a single host and a novel mechanism of genomic variation. *Mol Microbiol, 54*(2), 366-375.

Peterson, J. D., Umayam, L. A., Dickinson, T., Hickey, E. K., & White, O. (2001). The comprehensive microbial resource. *Nucleic Acids Res, 29*(1), 123-125.

Prangishvili, D. (2003). Evolutionary insights from studies on viruses of hyperthermophilic archaea. *Res Microbiol, 154*(4), 289-294.

Prangishvili, D., Arnold, H. P., Gotz, D., Ziese, U., Holz, I., Kristjansson, J. K., et al. (1999a). A novel virus family, the *rudiviridae*: Structure, virus-host interactions and genome variability of the *sulfolobus* viruses sirv1 and sirv2. *Genetics, 152*(4), 1387-1396.

Prangishvili, D., Arnold, H. P., Gotz, D., Ziese, U., Holz, I., Kristjansson, J. K., et al. (1999b). A novel virus family, the rudiviridae: Structure, virus-host interactions and genome variability of the sulfolobus viruses sirv1 and sirv2. *Genetics, 152*(4), 1387-1396.

Prangishvili, D., & Garrett, R. A. (2004). Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochem Soc Trans, 32*(Pt 2), 204-208.

Prangishvili, D., & Garrett, R. A. (2005). Viruses of hyperthermophilic crenarchaea. *Trends Microbiol, 13*(11), 535-542.

Prangishvili, D., Stedman, K., & Zillig, W. (2001). Viruses of the extremely thermophilic archaeon sulfolobus. *Trends in Microbiology, 9*(1), 39-43.

Rachel, R., Bettstetter, M., Hedlund, B. P., Haring, M., Kessler, A., Stetter, K. O., et al. (2002). Remarkable morphological diversity of viruses and virus-like particles in hot terrestrial environments. *Arch Virol, 147*(12), 2419-2429.

Reiter, W. D., & Palm, P. (1990). Identification and characterization of a defective ssv1 genome integrated into a trna gene in the archaebacterium sulfolobus sp. B12. *Molecular & General Genetics, 221*(1), 65-71.

Reiter, W. D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W., & Grampp, B. (1987a). Identification and characterization of the genes encoding three structural proteins of the sulfolobus virus-like particle ssv1. *Molecular & General Genetics, 206*(1), 144-153.

Reiter, W. D., Palm, P., Voos, W., Kaniecki, J., Grampp, B., Schulz, W., et al. (1987b). Putative promoter elements for the ribosomal rna genes of the thermoacidophilic archaebacterium sulfolobus sp. Strain b12. *Nucleic Acids Res, 15*(14), 5581-5595.

Reiter, W. D., Palm, P., & Zillig, W. (1988). Analysis of transcription in the archaebacterium sulfolobus indicates that archaebacterial promoters are homologous to eukaryotic pol ii promoters. *Nucleic Acids Res, 16*(1), 1-19.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Plainview, NY: Cold Springs Harbor Laboratory Press.

Schleper, C. (1993). *Genetische elemente und transformation des extrem thermophilen archaeons sulfolobus solfataricus.* Unpublished Ph.D., Ludwig-Maximillians Universität., Munich.

Schleper, C., Kubo, K., & Zillig, W. (1992). The particle ssv1 from the extremely thermophilic archaeon sulfolobus is a virus: Demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci U S A, 89*(16), 7645-7649.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., et al. (2001). The complete genome of the crenarchaeon sulfolobus solfataricus p2. *Proc Natl Acad Sci U S A, 98*(14), 7835-7840.

Snijders, A. P., Walther, J., Peter, S., Kinnman, I., de Vos, M. G., van de Werken, H. J., et al. (2006). Reconstruction of central carbon metabolism in sulfolobus solfataricus using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis. *Proteomics, 6*(5), 1518-1529.

Snyder, J. C., Spuhler, J., Wiedenheft, B., Roberto, F. F., Douglas, T., & Young, M. J. (2004). Effects of culturing on the population structure of a hyperthermophilic virus. *Microb Ecol, 48*(4), 561-566.

Snyder, J. C., Stedman, K., Rice, G., Wiedenheft, B., Spuhler, J., & Young, M. J. (2003). Viruses of hyperthermophilic archaea. *Res Microbiol, 154*(7), 474-482.

Soppa, J. (1999). Transcription initiation in archaea: Facts, factors and future aspects. *Mol Microbiol, 31*(5), 1295-1305.

Stedman, K., Schleper, C., Rumpf, E., & Zillig, W. (1999a). Genetic requirements for the function of the archaeal virus ssv1 in *sulfolobus solfataricus*: Construction and testing of viral shuttle vectors. *Genetics, 152*(4), 1397-1405.

Stedman, K. M., Schleper, C., Rumpf, E., & Zillig, W. (1999b). Genetic requirements for the function of the archaeal virus ssv1 in *sulfolobus solfataricus*: Construction and testing of viral shuttle vectors. *Genetics, 152*(4), 1397-1405.

Stedman, K. M., She, Q., Phan, H., Arnold, H. P., Holz, I., Garrett, R. A., et al. (2003). Relationships between fuselloviruses infecting the extremely thermophilic archaeon sulfolobus: Ssv1 and ssv2. *Res Microbiol, 154*(4), 295-302
.

Stetter, K. O. (1995). Microbial life in hyperthermal environments. *ASM News, 61*(6), 285-290.

Stewart, P. L., Burnett, R. M., Cyrklaff, M., & Fuller, S. D. (1991). Image reconstruction reveals the complex molecular organization of adenovirus. *Cell, 67*(1), 145-154.

Suttle, C. A. (1999). Do viruses control the oceans? *Natural History, 108*(1), 48-51.

Suttle, C. A. (2005). Viruses in the sea. *Nature, 437*(7057), 356-361.

Suttle, C. A., Chan, A. M., & Cottrell, M. T. (1991). Use of ultrafiltration to isolate viruses from seawater which are pathogens of marine phytoplankton. *Appl Environ Microbiol, 57*(3), 721-726.

To, C. M., Eisenstark, A., & Toreci, H. (1966). Structure of mutator phage mu1 of escherichia coli. *J Ultrastruct Res, 14*(5), 441-448.

Tolstrup, N., Sensen, C. W., Garrett, R. A., & Clausen, I. G. (2000). Two different and highly organized mechanisms of translation initiation in the archaeon sulfolobus solfataricus. *Extremophiles, 4*(3), 175-179.

Torrella, F., & Morita, R. Y. (1979). Evidence by electron micrographs for a high incidence of bacteriophage particles in the waters of yaquina bay, oregon: Ecological and taxonomical implications. *Appl Environ Microbiol, 37*(4), 774-778.

Tuomi, P., Torsvik, T., Heldal, M., & Bratbak, G. (1997). Bacterial population dynamics in a meromictic lake. *Appl Environ Microbiol, 63*(6), 2181-2188.

Van Heel, M. (1987). Angular reconstitution: A posteriori assignment of projection directions for 3d reconstruction. *Ultramicroscopy, 21*(2), 111-123.

Van Regenmortel, M. H. (1990). Virus species, a much overlooked but essential concept in virus classification. *Intervirology, 31*(5), 241-254.

Verhees, C. H., Kengen, S. W., Tuininga, J. E., Schut, G. J., Adams, M. W., De Vos, W. M., et al. (2003). The unique features of glycolytic pathways in archaea. *Biochem J, 375*(Pt 2), 231-246.

Vierke, G., Engelmann, A., Hebbeln, C., & Thomm, M. (2003). A novel archaeal transcriptional regulator of heat shock response. *J Biol Chem, 278*(1), 18-26.

Wang, F., Wang, P., Chen, M., & Xiao, X. (2004). Isolation of extremophiles with the detection and retrieval of shewanella strains in deep-sea sediments from the west pacific. *Extremophiles, 8*(2), 165-168.

Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., et al. (2003). The genome of nanoarchaeum equitans: Insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci U S A, 100*(22), 12984-12988.

Whitaker, R. J., Grogan, D. W., & Taylor, J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science, 301*(5635), 976-978.

Wiedenheft, B., Stedman, K., Roberto, F., Willits, D., Gleske, A. K., Zoeller, L., et al. (2004). Comparative genomic analysis of hyperthermophilic archaeal fuselloviridae viruses. *J Virol, 78*(4), 1954-1961.

Woese, C. R., Kandler, O., & Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Science USA, 87*, 4576-4579.

Wommack, K. E., & Colwell, R. R. (2000). Virioplankton: Viruses in aquatic ecosystems. *Microbiol Mol Biol Rev, 64*(1), 69-114.

Wommack, K. E., Hill, R. T., Muller, T. A., & Colwell, R. R. (1996). Effects of sunlight on bacteriophage viability and structure. *Appl Environ Microbiol, 62*(4), 1336-1341.

Yeats, S., McWilliam, P., & Zillig, W. (1982). A plasmid in the archaebacterium sulfolobus solfataricus. *The EMBO Journal, 1*(9), 1035-1038.

Zhang, Y., Strasser, P., Grabherr, R., & Van Etten, J. L. (1994). Hairpin loop structure at the termini of the chlorella virus pbcv-1 genome. *Virology, 202*(2), 1079-1082.

Zillig, W., Arnold, H. P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., et al. (1998). Genetic elements in the extremely thermophilic archaeon sulfolobus. *Extremophiles, 2*(3), 131-140.

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., et al. (1994). Screening for *sulfolobales*, their plasmids, and their viruses in icelandic solfataras. *Systematic and Applied Microbiology, 16*, 609-628.

Zillig, W., Prangishvilli, D., Schleper, C., Elferink, M., Holz, I., Albers, S., et al. (1996). Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic archaea. *FEMS Microbiol Rev, 18*(2-3), 225-236.

Zillig, w., Stetter, S. O., Wunder, S., Schulz, W., Priess, H., & Scholz, I. (1980). The *sulfolobus-"caldarialla"* group: Taxonomy on the basic of the structure of DNA-dependent rna polymerase. *Archives of Microbiology, 125*, 259-269.

APPENDIX A

CULTURE INDEPENDENT VIRUS SURVEY IN YNP

CULTURE INDEPENDENT VIRUS SURVEY IN YNP

Materials and Methods

As a first pass at isolating viruses from the thermal features in YNP we employed a technique that had been shown to be successful in marine environments. Using this method we concentrated large volumes of environmental sample by pumping it through a Pellicon 22 micron tangential flow filter (Millipore Corp.).

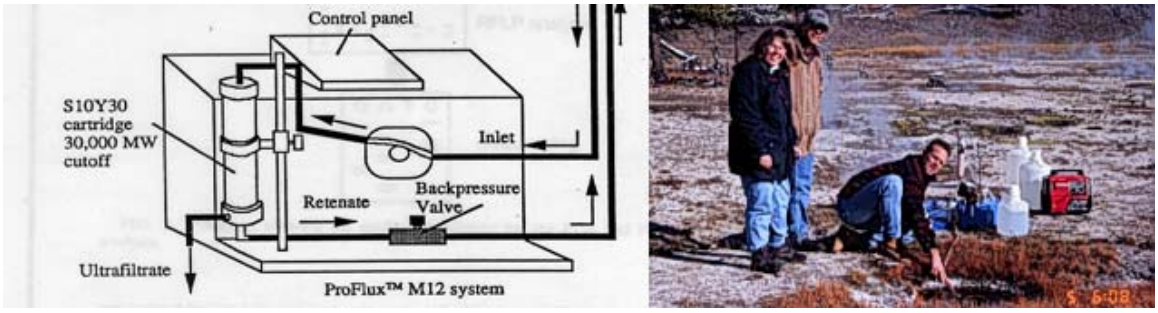

Figure A.1 - The schematic on the left shown an ultra-filtration apparatus (Millipore Corp.) similar to the one used in this study (shown on the right). In both filtration setups a large volume of water taken directly from the environment is drawn through a series of filters (glass wool down to .7 μm) in order to remove larger particulate that could clog the tangential flow filter. Sample that has passed the large particulate filters is then pumped through a .2 μm filter that removes particulate the size of most bacteria and larger. The liquid containing the remaining particles is then pumped over a tangential flow filter (30,000 – 100,000 MW cut off) fitted with a backpressure valve on the flow through port. The backpressure created by an adjustable valve on this flow through forces every thing smaller than the cutoff weight out the ultra-filtrate port. This ultra-filtrate waste and any particles that are virus size exit via the flow through port and are collected for further processing. This process can be repeated with the retained shrinking the total volume of retenate by the amount that passes out the ultra filtrate port as waste. This process can be repeated as many times as necessary to shrink the initial volume of the sample down to the desired amount.

Table A.1 - UTM grid positions, temperate, and pH of sampling sites.

Site Name	X Position	Y Position	Temp.	pH
Fountain Flats	513307.2	4934500.0	62-87	3.0-4.1
Midway Basin	514867.9	4933745.0	45-89	4.0-8.2

In Yellowstone, we selected a series of sampling locations (Table A.1) in the general vicinity of, and including Octopus Springs, in the Midway Geyser Basin. A short 5/8 inch clear plastic tube was placed directly into the effluent channel of a thermal feature, or near the side of a feature if there was no out flow. In order to pick up less particulate tubing was suspended slightly above the bottom of the pool or channel then passed through a peristaltic pump to either a 100,000 MW cut off tangential flow filter, or into a 20 L carboy. If the sample was pumped directly into the carboy, it acted as a settling reservoir, allowing larger particulate to settle out before being processed further. The alternative was to place a cartridge type filter inline after the tubing passed through the pump. A fine mesh screen or wrap of glass wool was placed at the end of the tube to filter out larger particulate. If the cartridge filter was placed between the pump and the initial extraction point the tubing had a tendency to collapse due to the negative pressure created between the two points. In these cases the less technical settling reservoir method was employed.

After pre-filtering the sample it was pumped through the Pellicon 0.22-micron tangential flow filter. Flow rate was controlled both by the rate of the peristaltic pump

and by placing guillotine clamp over the tubing on the flow through of the tangential flow filter thereby controlling the amount of fluid that passes through it. The more the flow through was constricted, the more fluid was forced through the filter. Particulate larger than 100,000 MW was not able to pass through the filter and was retained in the flow through (retenate) that is being constricted by the clamp. This tube was then placed in another 20 L carboy for collection and the clamp on the retenate was adjusted so the volumes of the flow through (retenate) and waste (ultra-filtrate) were approximately equal reducing the original volume by half. This process of re-filtering the total volume of the sample can be repeated over and over again with a 50% reduction in volume each time until the total sample is concentrated to the amount desired. We aimed to start with a total volume between 100-200 L and process to a final volume of less than a liter, thereby achieving a concentration 100-200 x the original sample.

All samples collected in this manner were then processed in one of four ways: PEG (polyethylene glycol) precipitate, further concentration through a 50ml Centricon filter, sedimentation by centrifuging at 26,000 rpm for two hours, or spotted on culture plates. Each of these approaches are detailed below.

The PEG precipitate method was preformed by adding 15% weight/volume PEG 8000 and 1molar NaCl to virus buffer (10 mM EDTA, 5 mM NaAC, and 50 mM NaCL at pH 4.8). The solution was stirred over night at 4°C and then spun for 20 minutes at12,000 x g. When the proteins (such as a virus capsid) are precipitated in this manner, it is often possible to collect them by centrifugation at 10,000 x g. The resulting pellets were re-suspended in 1 ml virus buffer and dialyzed against the buffer. The remaining

supernatant was centrifuged at 30k for 50 minutes to pellet virus and any virus pellet that formed was re-suspended in a minimal amount of virus buffer (usually 1/100th of the original volume).

The centricon flow-through is a way of further concentrating any virus size particles that may have been collected by further reducing the volume through another filtration step. The centricon unit is able to process smaller amounts (50 -5 ml) than the filters employed in previous treatment. This was performed by adding the processed sample to the self-contained filters approximately 40 mls at a time. The centricon filter is spun at 3.5 K until most of the liquid has passed through the filter. In this manner particles that are larger than 100 KD molecular weight are retained in the flow through. This process is repeated by reloading the same filter with more liquid until the total volume of the original sample processed. In this manner any virus size particulate from the total sample is retained on the filter and the total volume is reduced again to just a couple of mls. To speed the process two to three filters were usually used at the same time. After the entire sample has been processed retained particulate is washed of the filter by pipeting a small volume of remaining liquid (retenate) repeatedly over it. The filter was then inverted into a collection cap and spun again at 6,000 x g to collect the particles.

The third method consists of direct high-speed centrifugation. It involves filling up 26 ml centrifuge tubes with the pre-filtered sample and spinning them for 2-3 hours in a type 30 rotor at 25,000 x g. After sedimentation, the liquid is poured out of the tube gently so as not to dislodge any pellet that may have formed. Tubes were set upside

down on a paper towel to drain briefly and tapped lightly while still inverted to remove any remaining liquid. Any pelleted material was collected by pipeting a minimal amount of virus buffer or liquid from the original sample over a pre-marked spot on the tube where the pellet is most likely to form. It is necessary to pre-mark the centrifuge tubes because some time the pellets are transparent any very hard to detect visually.

The final method was to pour a thin film of sample (approximately 0.5 ml) onto standard one and a half percent agarose plates supplemented with 0.1 percent tryptone and incubated at 45°C. These plates were checked periodically for any type of colony formation. The principal of this approach being to look for the presence of any microbes that may be culturable in theses environments in a non-selective manner.

Each on of these preparations were then further assayed for virus using one or more of the following techniques described in detail below: Direct visualization with the TEM, nucleic acid extraction with the hot phenol method (Sambrook et al. 1989), run on a PHASTA gel to look for protein, Bio Rad protein assay, and PCR with virus specific primers.

Direct visualization with the TEM was performed by spotting 5 µl of sample on a Formvar and carbon coated copper EM grid and staining it with 2% urinal acetate for 30 seconds. Grids where then looked at with a Zeiss 100CA transmission electron microscope for presence of virus like particles. Numerous fields of view (grids) between amagnification of 5,000-50,000 were scanned for particles that had any regularity or symmetry in morphology. It was often apparent when scanning in a random manner at low magnification (5,000-8,000) if there was any particulate present necessitating a

closer look. If there were particles that had virus like qualities then as many grids as necessary were inspected at higher magnifications (25,000-50,000) to further determine their nature. If there were fields of view in which virus like particle were obvious in, then pictures were taken with the Zeiss 100CA for documentation. A concentration of purified cowpea chlorotic mottle virus (CCMV) was added in successive dilutions as a control.

Total nucleic acid extraction of concentrated environmental sample was attempted if there was a pellet produced from any of the methods discussed above. This was done with a standard phenol chloroform extraction method (Sambrook et al. 1989) with the following modifications. Briefly outlined, phenol was added to the sample and then heated for 30 minutes at 65°C. This was followed by vortexing the sample for approximately 30 seconds and spinning it at 14 K for 5 min. in a ss34 rotor. The upper aqueous phase and a tiny bit of the interface is pipetted off with the aqueous layer and a second equal volume of phenol/chloroform (24/1) was added and the procedure was repeated with out heating or drawing any of the interface while removing the aqueous portion for the second time. Nucleic acids were precipitated with two volumes of ethanol are added to the remaining layer with 0.1 x volume of 3 M NaOC and put on ice for 30 minutes to over night. This was then centrifuged at 14,000 x g for 20 minutes and the liquid carefully drawn off the pellet with a pipette. The pellet was then washed with an 85% ethanol solution, and spun again for 5 minutes. Finally the liquid is drawn off the pellet again and it is left open or placed in a vacuum oven to dry. The pellet is then re-

suspended in 10 mM Tris 1 mM EDTA (TE) and run on an 1% TBE agarose- gel (Maniatis). CCMV was added in successive dilutions to use as a control.

Collected environmental samples were assayed for the presence of protein. Total proteins were extracted by boiling the samples in laemmli disruption buffer (50 mM Tris pH 6.8, 2% SDS, 8% BME) for five minutes followed by cooling on ice for at least three minutes. One μl of the denatured sample was run on an SDS-PAGE (sodium dodeycl sulfate-polyacrylamide gel electrophoresis) denaturing gel and stained with silver stain. Specifically, an Amersham Pharmacia (now GE Healthcare, www.4amershambiosciences.com) Phast System® integrated gel and stain system was used to run 12.5% homogeneous Phasta gels with SDS buffer strips followed by silver stain (Sambrook et al., 1989). CCMV was added in successive dilutions to use as a control.

The Bio-Rad Protein Assay kit (Bio-Rad Laboratories Inc., www.bio-rad.com) was also employed to look for the presence of protein in the processed samples. This is a method for determining the concentration of solubleized proteins. This is a dye-binding assay in which a differential color change of a dye happens in response increasing concentrations of protein. Measurement of dyed samples compared to a standardized curve with a spectrophotometer at 595 nm provides a relative concentration of the protein being assayed. A standard curve was created by measuring the absorbance (at 595 nm) of successive dilutions of bovine serum albumin (BSA). CCMV was added in successive dilutions to use as a control.

PCR Primers were designed from ORF's in the SSV1 virus to be used to search for viruses that may be similar in YNP's thermal acidic sites.

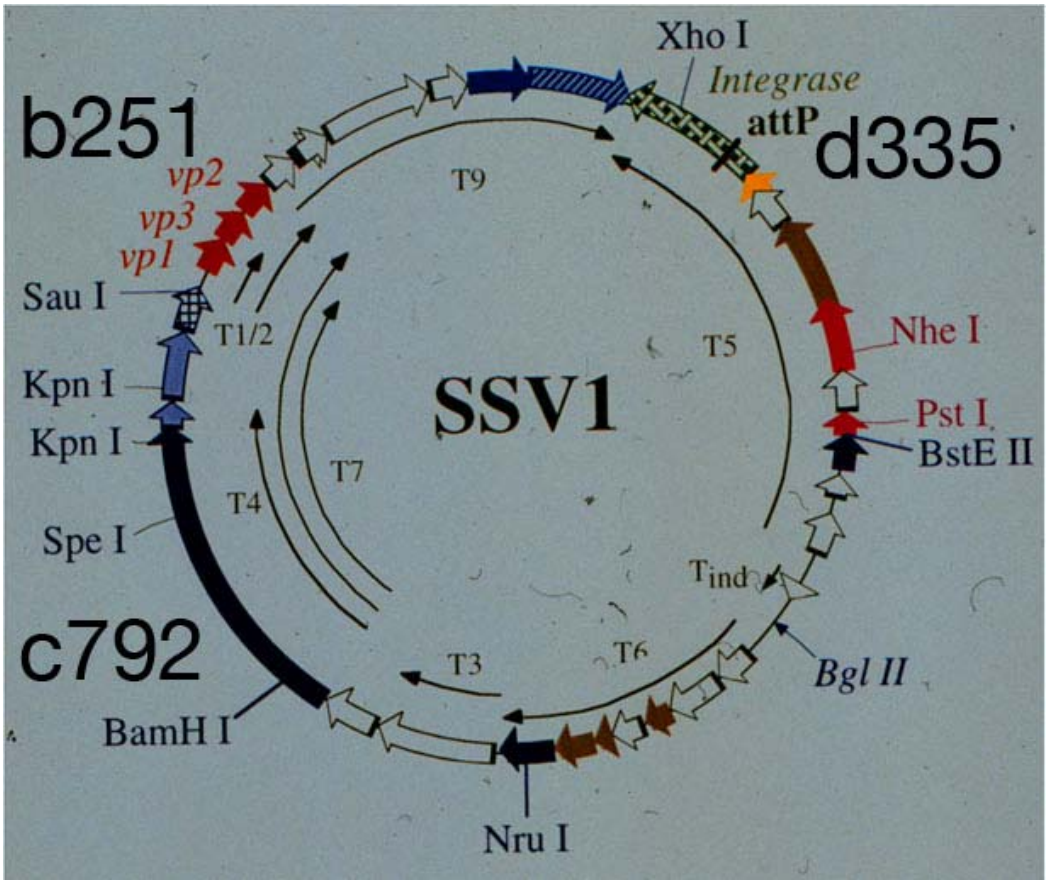

Figure A.2 - (K. Stedman et al. 1999a) Diagram of the SSV1 circular genome with the open reading frames denoted by the bold arrows on the circle. Areas of the genome that were used to design primers to look for new thermal viruses are marked in larger type black letters and numbers.

Three sets of primers were designed from three separate ORFs in the SSV1 genome that we reasoned would be more likely to be conserved, over space and time, than other regions of the genome. These primers were named SSV1-1 forward and reverse (F & R), SSV1-2 (F & R), and SSV1-3 (F & R). The SSV1-1 primer set was designed to

amplify the b251 ORF that codes for the viral coat proteins in SSV1 (SSV1-1F is AGCCCCATAAGACTAACGGC, SSV1-1R is GTAGTGATGGGTTTGACGGG). SSV1-2 primers were designed for the d335 of SSV1, a region that codes for the viral integrase gene (SSV1-2F is CCCCTTTTAGCCATTCCGC, SSV1-2R is ACGCGAGAGGAAAGGGCGG). SSV1-3 primer set amplifies the c792 ORF in SSV1, which is the largest ORF in the virus's genome (SSV1-3F is CGGCGGAGGAGGGCCCAAC, SSV1-3R is CCTCCCTCCCCATGAAACGG).

Results

Controls were run to determine sensitivity on the above methods with CCMV and BSA. Specifically I looked at sensitivity levels of the phasta gels, how and if filters were effecting detection and if centrifugation was a viable method of pelleting virus sized proteins. The TEM was also used to correlate detection levels of CCMV with visible amounts of virus on the EM grid.

We were able to detect down to 20 ng/μl of CCMV run on phasta gel after 50 mls of the virus diluted to that concentration was run through a centricon 100 and prepared for the gel. We could also detect 100 ng/μl of both BSA and CCMV with the Bradford Assay. Aliquots were taken from controls prepared before spinning at 36 K for 2 hours and after re-suspension of the pellet.

I was able to detect large bands with a smear the entire length of Phasta gel (Figure A.4) after loading 1 ug of CCMV spun at 38,000 x g for 6 hours and re-suspending the pellet in 250 μl of virus buffer. There was nearly solid virus on grids

evident when checked with the TEM in this sample, as one would expect with a virus concentration of that magnitude.

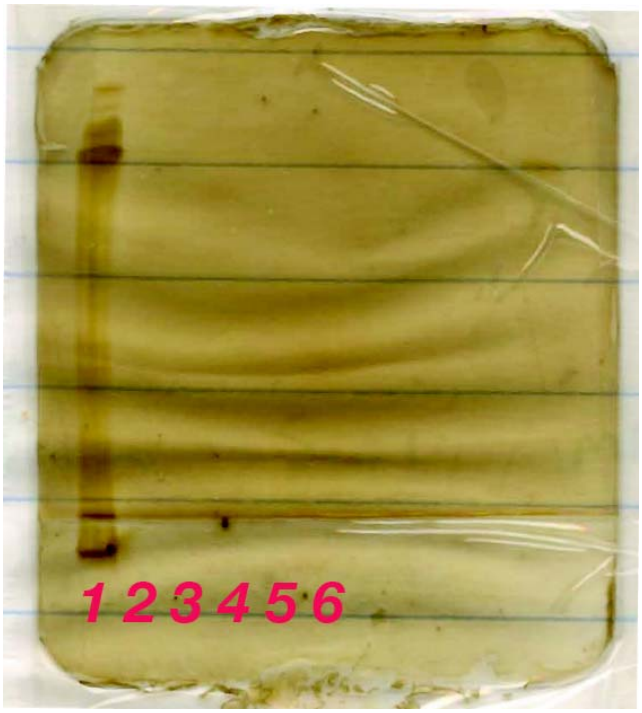

Figure A.3 - Phasta gel with 1 mg of CCMV treated in the fashion described above. Lanes two through six are from samples that were centrifuged and then the pellets re-suspended in 250 µl of virus buffer. Lanes two through four were the concentrate from a 20 l filtered in the field and then treated with 10% PEG and 1 M NaCl. Lane two was unfiltered, lane three was filtered, and lane four was the supernatant after a low speed spin. Lane five and six were re-suspensions of the pellets from lanes four and five.

In another control designed to determine if protein was being lost in filtration we could see down to 10 ng of BSA before and after running through a .2 µm filter on the phasta gel (Figure A.4).

Figure A.4 - Phasta gel showing the effects of filtration upon protein levels run with BSA. In lanes one through four the protein was pushed a 0.22 μm filter and in lanes five through eight the protein is not processed in any way. From lane one to four the concentration of BSA rises from 100 ng, to 1μg, 10 μg, and 100μg consecutively. In lane five through eight the protein concentration increases in the same manner.

Discussion

It is probable that in these extreme thermal, mostly acidic environments that were investigated a combination of two factors makes the detection of virus by direct filtering and concentration of environmental sample difficult. The first factor is that the amount of biomass present in many of the thermal features is most likely orders of magnitude below that which is found in the trophic zone in marine environments where this methodology has been successful. The second is that most of the viruses that may be

present are not lytic in nature. In retrospect, neither of these factors surprising insofar as one might expect to find greatly reduced biomass in such an extremely harsh habitat. Additionally from a viral perspective it would be advantageous to use the host (by remaining intra-cellular) either in an assembled form as an extra-chromosomal element or by incorporating into the host genome emerging only in order to propagate your own genome. Both of these factors would conspire to make it extremely difficult to detect viruses independent of their hosts present in these environments. It is also possible that in some of the thermal features we investigated, especially the muddier ones if there were virus particles present they could be masked. If there were clays present in sufficient quantities and the viral particles were prone to adhesion to these minerals it would make detection exceedingly difficult.

The controls that were run with CCMV and BSA would indicate that the detection levels of the centrifugation filtering and phasta gels, would fail to indicate the presence of virus in these mainly oligotrophic environments. Had everything been working perfectly and had there been free virus present in the amount found in marine trophic environments it is possible we would have detected them as such. It is possible that we did not sample the thermal features with the right combination of methods, due to our experience level at that point, to find VLP's.

Since this time others in the lab have had better results with similar approaches. These include Epifluorescence microscopy (EFM) of YNP acidic hot springs that estimate cell densities that vary between 10^5-10^6 cells/ml (Ortmann et al. 2006). This is still well below 10^8 cell/ml one can experience in marine environments but quantifies an

amount that approaches levels of biomass that if producing virus would be in the range of detectable levels.

Direct isolation of VLPs from hot springs has also been accomplished. Epifluorescence microscopy estimates of free virus particles in acidic hot springs range from 10^3-10^5 VLPs/ml (unpublished data), which are orders of magnitude lower than those observed in marine environments. This reduced number of VLPs may reflect the adaptation of viral lifestyles that minimize the exposure of virus particles to the harsh environment. Direct filtration of hot spring water, selecting for VLP-sized components, have been successful for members of the Young lab (*unpublished results*), but the discussion of this progress is beyond the experience of this author.

REFERENCES CITED

Ortmann, A. C., Wiedenheft, B., Douglas, T., & Young, M. (2006). Hot crenarchaeal viruses reveal deep evolutionary connections. *Nat Rev Microbiol, 4*(7), 520-528.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Plainview, NY: Cold Springs Harbor Laboratory Press.

Stedman, K., Schleper, C., Rumpf, E., & Zillig, W. (1999). Genetic requirements for the function of the archaeal virus ssv1 in *sulfolobus solfataricus*: Construction and testing of viral shuttle vectors. *Genetics, 152*(4), 1397-1405.